



The Alberta Journal of Educational Research

Vol. XIV, No. 2

June, 1968

CONTENTS

- Multivariate Analyses of Test Responses as a Pre-Requisite to Item Analysis 95
Ram K. Gupta
- A General Theory of Innovativeness 101
H. David Hemphill
- Assessing Intellectual Potential of Native Canadian Pupils: A Summary..... 115
R. S. MacArthur
- Comparison of a Disguised and Reported Measure of Risk
Taking on Objective Examinations 123
Malcolm J. Slakter and Aubrey H. Roden
- The Use of Expectancy Tables in Educational Guidance 129
Dale G. Andersen and Herbert E. Stiles
- Transformations: Length Conservation 137
Daiyo Sawada and L. Doyal Nelson

BOOK REVIEWS

- David A. Payne, *The Specification and Measurement of Learning Outcomes*, (Waltham, Massachusetts: Blaisdell, 1968), V. R. Nyberg 151

COMMITTEE ON EDUCATIONAL RESEARCH
The University of Alberta



The Alberta Journal of Educational Research

A SPECIAL ISSUE THE USE OF COMPUTERS IN EDUCATION

A special issue on The Use of Computers in Education is being planned for 1969. Several manuscripts are in hand but scholars interested in this area are invited to submit manuscripts or give notice of intention to do so.

Editorial Comment

The funding of research in Canada is a problem of major proportions. It is particularly difficult to get funds in sufficient quantity to research problems in education. New agencies, such as the Human Resources Research Council in Alberta, have been created to fund research into social problems, including education. Will these agencies be able to provide enough money to finance all of the research necessary in these areas? Perhaps more important, will the agencies be able to fund the projects that qualify as basic-research as well as the developmental projects?

The term "basic-research" is difficult to define. It has been variously described as "when you don't know what you're doing," as "the opposite of practical research," and as "farfetched and useless." It could be said that basic-research is a creative synthesis; it is not seeing but discovering. However, it is probably not as important to define basic-research as to be able to differentiate between basic-research projects of greater and lesser importance.

The greatest need in any field of research endeavour is to recognize significant basic-research in its earliest stages before any practical applicability is apparent. There is the need to convince those charged with dispensing the funds that the research is important. Research based on deductive reasoning, more properly called development, is relatively easy to finance because of its immediate applicability to everyday problems which can be described precisely in a routine application for funds. Those involved in basic-research have the problem that original or creative research cannot be planned in advance. Because basic-research is discovery, it cannot be anticipated wholly on the basis of known facts and generalizations.

Researchers investigating fundamental problems in education must accept responsibility for translating their problems into a language meaningful to the layman. The layman, in turn, must realize that, however simplified, the essence of basic-research cannot be assimilated without effort on his part. Can educational researchers give assistance to funding agencies in establishing guides for determining which are the better basic-research projects? Selye¹ uses three criteria to identify significant scientific *discoveries*. Is it possible to restate these criteria so that they might form a useful basis for making decisions about projects in their early stages? Selye states his criteria in this way:

. . . it is characteristic of great basic discoveries that they possess, to a high degree and simultaneously, three qualities: they are true not merely as facts but also in the way they are interpreted, they are generalizable and they are surprising in the light of what was known at the time of the discovery [p. 149].

The importance of the word *simultaneous* must be emphasized. Selye illustrates, from the field of medical research, discoveries that would

¹ See an article by Hans Selye, a prominent medical researcher, in Thruelsen, R. T. and Kobler, J. (Eds.). *Adventures of the mind*, New York: Vintage Books, 1960.

satisfy one or two of the criteria, but not all three. These would not, in his opinion, be classified among the great discoveries.

It is not sufficient for researchers to complain about the lack of funds for educational research, particularly for projects that would be classified as basic-research. They must be ready to help in stating criteria by which funding agencies can identify the projects most worthy of support.

* * *

A special issue of the *Alberta Journal of Educational Research* is being planned for 1969. As in the issues devoted to models in education and to the work of Piaget, this special issue will be on one theme: The Use of Computers in Education. Several manuscripts on this theme are already on hand; others will also be considered for inclusion in the special issue.

Articles of two types seem pertinent to the special issue. One type would be broadly based, setting the stage for an insightful look at the total area of computers and education. The second type would include articles on the use of computers in the research process in education or in the instructional process, including evaluation. Interested scholars are invited to submit manuscripts related to this theme, or to give notice of their intention to do so.

* * *

In this issue of the *Journal*, Gupta and Hemphill explore the theoretical aspects of two diverse educational questions. Gupta discusses the appropriateness of traditional methods of item analysis and presents theoretical support for the use of multivariate techniques for analyzing item responses. Hemphill argues for the need to look at innovativeness within the larger framework of general systems theory.

MacArthur presents a synthesis of several projects investigating methods of assessing the intellectual potential of Canadian native pupils. He summarizes data obtained in a series of longitudinal studies carried out under his direction.

Andersen and Stiles report a study of the use of expectancy tables in advising students in small rural high schools. Counsellors will have to weigh the convenience of such an approach against its use by untrained personnel.

The article by Roden and Slakter reports an empirical study of the effect of risk taking on student scores on multiple-choice examinations. They compare the student's perception of the risk he takes, and his willingness to do so, with the statistical evidence of his tendency toward risk taking.

In the last issue of the *Journal*, Sawada and Nelson presented an argument for a method of assessing length conservation that was independent of verbal facility. In this issue they present the results of a study utilizing this method. Their results are compared with the results of other researchers using a verbally-dependent method to determine conservation.

P.A.L.

The author presents empirical evidence derived from cluster and factor analyses of item responses in support of the argument that for developing internally consistent or homogeneous measuring instruments, multivariate techniques for analyzing responses should get precedence over the traditional methods of item analyses.

R A M K . G U P T A
The University of Alberta

Multivariate Analyses of Test Responses as a Pre-Requisite to Item Analysis

Problem

Item analysis of one type or the other is universally regarded as an indispensable step in developing objective tests. An important objective of item analysis is to ensure internal consistency (also called homogeneity and coefficient alpha) so that the test may consistently measure "whatever it purports to measure" or "whatever it really measures." The phrases within quotes refer respectively to the claimed and the actual *validities* of the test. Ideally, criterion measures should be obtained independent of the test being developed which may be intended to be a more efficient substitute for the procedure leading to the criterion measures. Since external criteria having acceptable reliability are rarely obtainable, recourse is generally taken to an internal criterion—the scores on the test itself, the items of which are being analyzed.

The present research poses two questions. The first relates to item analysis against criteria in general—external or internal—and can be stated as: "Is it possible to develop a homogeneous test if item analysis is done, using a heterogeneous criterion?" Logically, the answer would seem to be in the negative.

The second question is specifically related to the use of the scores on

the test itself as an internal criterion which is by far the commonest criterion used for item analysis. Since the test scores are based on the responses to the individual items, accepting these scores as an internal criterion should logically imply that the individual items also are acceptable, in which case item analysis would be superfluous. On the other hand, a lack of confidence in the adequacy of the items would imply that the test scores also constitute an inadequate criterion against which item analysis would be neither desirable nor justifiable. Is there a way out of this dilemma or vicious circle?

The present research seeks an answer to the above questions simultaneously insofar as they relate to the use of internal criteria for item analysis.

The problem as posed above points to the desirability of making the criterion homogeneous before using it for item analysis. When the criterion is internal, its homogeneity can be examined easily through multivariate analyses of the responses to the items constituting the criterion.

Data and Procedures

The data of this research consisted of the responses given by 460 girls and 554 boys to the fifty items of the Sequential Tests of Educational Progress, Mathematics, Form 2A (STEP-Math. 2A). The subjects were students of ninth grade mathematics in twenty-nine schools in the U.S. mid-west.

The reliability of the test in terms of Kuder-Richardson formula twenty (K-R-20) was found to be .71 for girls and .75 for boys. In view of the facts that (a) the subjects were substantially heterogeneous, (b) the instrument was a widely used standardized test, and (c) the nature of the subject matter would normally lead to high reliabilities, the obtained figures seemed rather low. It was decided, therefore, to examine the test for homogeneity before the test scores were used for item analysis. Five different multivariate analyses were used for this purpose. Each was applied separately to the responses of the boys and those of the girls. They were:

1. The method of "homogeneous keying" of DuBois, Loevinger, and Gleser (1952), applied to the inter-item matrices of variances and covariances,
2. principal factor analysis of phi-coefficients,
3. principal factor analysis of tetrachorics,
4. alpha factor analysis of phi-coefficients, and
5. alpha factor analysis of tetrachorics.

In all the factor analyses, squared multiple correlations were used for communalities wherever possible, four factors were extracted, and the resulting common factor space was subjected to normal varimax rotation.

The rationale for selecting the above methods of multivariate analyses and a detailed comparison of the results given by them forms the subject matter of another paper now under preparation. It is sufficient to state here that each of them revealed that the test was undoubtedly heterogeneous, containing two homogeneous subgroups of items representing two independent abilities, leaving outside their fold about a third of the fifty items.

- The above was followed by steps 1 through 4 given below:
1. Using the scores on *all the items* of the test as the internal criterion, biserial correlations were calculated for each of the fifty items;
 2. Using the scores obtained *from the items falling in a given factor* as the internal criterion, biserial correlations were found for only those items which were included in the factor;
 3. Calculations corresponding to 2 above were done using scores on each of the clusters as the internal criterion;
 4. Calculations of K-R-20 reliabilities were done for the results of analyses 2 and 3.

Results and Conclusions

Table 1 shows the biserial correlations and K-R-20 reliabilities for boys' responses when all the items constituted a single internal criterion and also when the biserial correlation of an item was calculated with reference to that segment of the internal criterion for which the item had shown maximum affinity, considering the results of multivariate analyses. Table 2 contains the corresponding information for girls.

The results of Tables 1 and 2 led to the following conclusions:

1. Factor analyses split the test into three parts containing 17, 19, and 14 items in the case of boys and 15, 18, and 17 in that of girls. Similarly, the method of "homogeneous keying" split the test into three parts, containing 14, 17, and 19 items in the case of boys and 15, 18, and 17 in that of girls (Vide second last row in Tables 1 and 2). The first two factors and clusters were highly homogeneous; the third was extremely heterogeneous, as evidenced from the reliabilities given in the last row of the tables. In fact, the heterogeneous part merely contained the items not included in the homogeneous clusters or factors.

TABLE 1
BISERIAL CORRELATIONS AND K-R-20 RELIABILITIES FOR
BOYS' RESPONSES TO STEP-MATH 2A
CLASSIFIED BY NATURE OF CRITERION SCORES

Item no.	Total test	Nature of Criterion Scores					
		Fact. I	Fact. II	Fact. III	Clus. I	Clus. II	Clus. III
1	-.053			.332			.192
2	.363	.510					.358
3	.500	.715			.720		

Item no.	Total test	Fact. I	Fact. II	Fact. III	Clus. I	Clus. II	Clus. III
4	.205		.503			.493	
5	.464	.715			.725		
6	.310			.479			.465
7	.493	.660			.677		
8	.632	.766			.784		
9	.641	.889			.897		
10	.618	.860			.867		
11	.515	.706			.726		
12	.448	.627			.629		
13	.506		.656			.680	
14	.318	.505					.447
15	.390			.375			.316
16	.653	.942			.962		
17	.600	.870			.876		
18	.576	.823			.844		
19	.386	.508					.185
20	.259			.424			.402
21	.474	.635	.475		.648		
22	.518					.487	
23	.521	.688			.697		
24	.551	.754			.748		
25	.231		.434				.231
26	.263		.398				.316
27	.338		.531			.538	
28	.219			.504			.419
29	.169			.417			.360
30	.295			.255		.471	
31	.411		.564			.578	
32	.259		.435				.340
33	.229			.390			.338
34	.478		.725			.744	
35	.338		.552			.549	
36	.155			.393			.311
37	.277			.412			.353
38	.363		.584			.568	
39	.395		.566			.580	
40	.394		.627			.632	
41	.434		.626			.645	
42	.204			.433			.353
43	.290		.499			.503	
44	.485		.656			.688	
45	.554		.745			.757	
46	.007			.333			.266
47	.465		.639			.649	
48	.301		.421				.370
49	.257			.338		.456	
50	.121			.369			.347
No. of items	50	17	19	14	14	17	19
K-R-20	.749	.808	.705	.175	.801	.703	.161

2. Item biserials improved substantially when the internal criterion was limited to the scores on the homogeneous subset of items as compared to the case in which the criterion scores were based on the entire test.

3. The Spearman-Brown formula worked in the reverse direction. For example, the K-R-20 reliability for the entire test of fifty items was .75 for the boys. When the number of items was reduced to only a third

TABLE 2
 BISERIAL CORRELATIONS AND K-R-20 RELIABILITIES FOR
 GIRLS' RESPONSES TO STEP-MATH 2A
 CLASSIFIED BY NATURE OF CRITERION SCORES

Item no.	Total test	Nature of Criterion Scores					
		Fact. I	Fact. II	Fact. III	Clus. I	Clus. II	Clus. III
1	-.007		.361				.343
2	.251			.439			.438
3	.457	.763			.762		
4	.287		.482			.480	
5	.372			.394			.395
6	.162			.320			.353
7	.389	.636			.642		
8	.479	.569			.588		
9	.466	.697			.688		
10	.478	.769			.777		
11	.469	.649			.640		
12	.280			.436			.462
13	.417		.555			.560	
14	-.010			.378			.448
15	.381	.550			.553		
16	.509	.897			.889		
17	.552	.831			.832		
18	.476	.782			.782		
19	.305	.555			.552		
20	.171			.314			.367
21	.347	.542			.539		
22	.508	.494			.600		
23	.302	.507			.495		
24	.438	.723			.729		
25	-.049			.266			.276
26	.247		.351				.346
27	.327			.377			.338
28	.280			.395			.387
29	.215			.323			.351
30	.401			.393		.446	
31	.471		.478			.579	
32	.441		.555			.553	
33	.332			.480			.444
34	.573		.629			.627	
35	.395		.517			.515	
36	.158			.309			.309
37	.460		.570			.565	
38	.363		.497			.498	
39	.483		.656			.650	
40	.417		.569			.563	
41	.383		.571			.572	
42	.211			.360			.351
43	.382		.518			.506	
44	.460		.615			.612	
45	.445		.642			.632	
46	.150			.353			.320
47	.446		.631			.621	
48	.360			.441		.446	.380
49	.283		.384			.386	
50	.209			.366			
No. of items	50	15	18	17	15	18	17
K-R-20	.709	.739	.674	.247	.745	.680	.236

of the total as in Factor I, the reliability increased to .81 though it should be dropped to .50.

4. The items which should have been rejected if analyzed against the scores on the entire test constituting the internal criterion seemed to be the ones which were essentially heterogeneous, for instance, items 1, 29, 36, 46, and 50 for boys; and items 1, 6, 14, 20, 25, 36, and 46 for girls. These items belonged to Factor III or Cluster III for which K-R-20 reliabilities had ranged from .16 to .25 only. The test resulting from the exclusion of these items still would have contained two strong, orthogonal factors or clusters and, in this sense, would have been heterogeneous. The only way in which the occurrence of more than one factor in a test can be revealed is the use of multivariate analyses.

5. Factor I and Cluster I contained almost the same items, the same applied to Factor II and Cluster II, and also Factor III and Cluster III. Thus, the choice of a particular multivariate method for analyzing matrices did not seem to be crucial.

Implications

1. Before item analysis is undertaken, it seems advisable to factor analyze the test items, especially if the test being developed is intended for wide use. However, should factor analysis not be possible in a given situation, the test developer might instead obtain the inter-item variance-covariance matrix and analyze it on a desk calculator, using the simple procedure of DuBois, Loevinger, and Gleser (1952). If the matrix contains a large number of negative and fairly large-sized covariances, the presence of more than one factor would be indicated.

2. When evidence suggests the presence of two or more strong, orthogonal factors, the test should be split into appropriate subtests, and the scores on each subtest should be used for analyzing the items included in the given subtest. Also, practical decisions should be based on subtest scores rather than on the scores for the entire test.

Reference

- DuBois, P. H., Loevinger, J., and Gleser, G. C. *The construction of homogeneous keys for a biographical inventory*. Air Training Command, Human Resources Research Center, Research Bulletin 52-18; Lackland Air Force Base, San Antonio, Texas, May, 1952.

This article presents generalizations on the adoption and diffusion of innovations. Using a general system framework, the author states these generalizations as theoretical propositions and maintains that they are applicable to systems which range in complexity from individuals to cultures. He suggests that the theory can encompass many of the theories of change and requires empirical investigation.

H. DAVID HEMPHILL

The University of Alberta

A General Theory of Innovativeness

If there is one truth most social theorists agree on and on which they can arrive at a quick consensus, it is the lack of a viable theory of social change (Bennis, 1966, p. 99).

Despite this lack of theory, however, there has been a great deal of research on the adoption and diffusion of innovations.¹ Unfortunately, few attempts have been made to synthesize the results of this research into a set of "universally applicable generalizations."² The purpose of this paper is to draw from empirical evidence and place in a general system framework a *general theory of innovativeness*.

The term "general" has been chosen because of the several senses in which it applies to the theory to be developed. First, the theory is general in the sense that the concepts used are sufficiently broad to encompass a wide range of specific phenomena. Second, the theory has a foundation in empirical evidence from diverse research areas, and is applicable to systems ranging in complexity from single individuals to

¹ For example, Ross cites over 150 studies on the adaptability of school systems (in *Administration for adaptability*. New York: Metropolitan School Study Council, 1958); and Rogers lists more than 500 references from a number of research areas (in *Diffusion of innovations*. New York: Free Press of Glencoe, 1962).

² Most notable among the recent attempts are those of Rogers (*Ibid.*, pp. 300-316); and D. E. Griffiths (Administrative theory and change in organizations. In Matthew B. Miles (Ed.), *Innovation in education*. New York: Bureau of Publications, Teachers' College, Columbia University, 1964. Pp. 425-436).

cultures.³ And finally, the propositions to be stated are general in the sense that they assume that all other factors impinging on the phenomena in question are controlled either experimentally or statistically.

The General System Model

Although general system theory may not be the *sine qua non* of current behavioural theory, as some believe it to be, it does provide a useful framework and lexicon of concepts upon which to base the theory to be developed.

Von Bertalanffy (1956, p. 3) defines a *system* as a set of elements standing in interaction. Hall and Fagen (1956) expand this to define a system as a set of elements together with the relationships between the elements and their attributes. It would appear that the simpler definition is satisfactory, however, because, in the final analysis, an element is not more than a set of attributes.

All but the smallest systems have *subsystems*, and all but the largest have *suprasystems*. The environment may be defined as the set of all elements which a change of the system will affect, and in which a change will affect the system (Hall and Fagen, 1956). A system may be open or closed. *Open systems* are in continuous interaction with their environment. That is, there is a continuous flow of *input* and *output* between the system and its environment. The system-environment interaction may be in the form of exchange of material or non-material resources, such as goods, services, information or energy. Open systems tend to maintain themselves in *steady states*. A steady state is a balanced relationship among the elements of a system and its environment.

One additional concept is vital to the understanding of system theory, this is the concept of feedback. *Feedback* is the process whereby some of the output of the system is fed back into the system as a guide for the system's further functioning. This paper will use the term *negative feedback* to mean that feedback which maintains stability of the system by exerting pressure on the system to return to its original state after a slight change. *Positive feedback*, on the other hand, is pressure on the system for continued change. Miller (1965) confirms this notion when he says, "positive feedback produces continuous increments of outputs which give rise to 'spiral effects' destroying one or more equilibria of a system [p. 402]."

In order that the general system model may be applied to behavioural phenomena, a number of assumptions must be accepted. First, it must

³ Boulding discusses a theoretical hierarchy of systems in terms of their complexity (in *General systems theory—The skeleton of science*. In Von Bertalanffy and Rapoport (Eds.), *General systems*. Yearbook of the Society for the Advancement of General Systems Theory: Vol. I, 1956. Pp. 11-17); and Miller develops hypotheses which span these levels of complexity (in *Living systems: Crosslevel hypotheses*. *Behavioral Science*, 1965 (October), Vol. X, 380-411).

be assumed that behavioural systems at all levels of complexity are open. Second, one must assume that these open systems maintain a steady-state, or an equilibrium, with their environment. Finally, it must be assumed that there are feedback structures within the system's environment.

Other Concepts

In addition to the concepts of the general system model, several other concepts specific to this theory must be defined.

First of all, it should be noted that *innovations* are a subset of the set of *change* phenomena. Barnett (1953) defines an innovation "as any thought, behavior, or thing that is new because it is qualitatively different from existing forms [p. 7]." Further, in this definition of innovation:

emphasis is placed upon reorganization rather than upon quantitative variation as the criterion of a novelty (innovation). Innovation does not result from the addition or subtraction of parts. It takes place only when there is a recombination of them [p. 9].

Innovations themselves may be further divided into internal and external innovations. Internal innovations are syntheses of existing thoughts and ideas into new forms within the system whereas external innovations are those generated in the system's environment. The theory to be presented deals primarily with external innovations.

Adoption of an innovation by a system is the acceptance and implementation of the innovation within the system. Thus, *innovativeness* will be defined as the tendency of a system to adopt a given innovation or set of innovations. This tendency has been measured in terms of whether or not the system adopts, to what extent it adopts, or how early it adopts relative to other similar systems in its environment.

Finally, the concept of system *structure* must be defined. For the purposes of this discussion, structure will be defined as the demonstrated, recurrent pattern of behaviour of the system. Thus, at the individual level, the person's structure will consist of such elements as his values, habits, attitudes, beliefs, needs and goals, together with the relationships among them. The structure of a group will include an aggregate of the individual's structures together with the group's norms, roles, status hierarchy, communications networks, power and authority relations. The organization structure will be similar to that of a group with the addition of elements such as formal task and incentive structures or what Andrews and Greenfield (1966-67) call "organizational themes." At the cultural level, the structure would include such things as folkways, mores, customs, ethos, values, and ideology;

It should be noted that the structure of a system, as defined here, is a set of the system's subsystems together with their relationships. These relationships thus express dependence or independence, and subordinate

or superordinate structures of the subsystems. An example of system structure at the individual level would be a hierarchy of needs as hypothesized by Maslow (Lazarus, 1963).

Thus, the perspective to be taken is that innovativeness is related to the characteristics of the system and its interaction with the environment, and not that innovativeness is an element of the system itself, as suggested by Miles (1965).

Theoretical Propositions

With the general and specific concepts related to this theory defined, it is now possible to state a number of propositions. Following each proposition, a brief discussion of supportive research findings will be presented. It should also be noted here that the generality of the propositions necessitates that they be viewed in terms of *Ceteris Paribus*. That is, rather than compound the statements of the propositions, they should be read as if they were preceded by the phrase "all other things being equal."

Proposition 1. The greater the input of information about innovations to a system from its environment, the higher its innovativeness.

It would appear obvious that a system cannot adopt an innovation from its environment unless it has information about that innovation. Therefore, it seems reasonable to expect innovativeness to vary directly with the amount of information input.

This proposition finds support in the findings of a number of studies. At the individual level, research among farmers (Lionberger, 1960), physicians (Coleman, Katz, and Menzel, 1957), and school superintendents (Carlson, 1965) indicates that innovativeness is positively related to amount of formal education, subscription to research journals, membership in professional organizations, and visits to urban centres or professional conferences.

Further support can be found in psychological research. Using Reisman's (1961) typology of inner-versus other-directed individuals, Linton and Sofer (1959) found other-directed people were more likely to change than inner-directed ones.

At the organizational level, innovativeness has been shown as positively related to deliberate surveys of new ideas, willingness to look outside the firm, connection with universities or research institutes, and hiring of "outsiders." These relationships have been established for many types of business, industrial (Carter and Williams, 1957), and educational organizations. For example, Carlson (1962) says:

... it has been demonstrated that when a school system calls for an insider, it calls for executive performance bent on system maintenance, and that when a school system calls for an outsider, it calls for executive performance bent on changing the system [p. 69].

Finally, it would not seem unreasonable to suggest that one of the main reasons for the acceleration of change in contemporary culture lies in technological advances in communication which have vastly increased the input of information to our cultural systems. Moore (1959) implies this when he suggests that "the greater the number of lines of contact, the greater the force of new influence [p. 179]."

Each of the variables which has been mentioned above may be interpreted as a measure of the quantity of input of innovation information to the system. Thus, they tend to confirm Proposition 1.

Proposition 2. The greater the perceived compatability of the innovation(s) with system structure, the higher the system innovativeness.

The first proposition dealt with the quantity of information input, whereas this one deals with the perceived quality of the innovation(s).

This proposition also enjoys the support of empirical evidence in a number of fields. For example, Fliegel (1956) found a significant relationship between innovativeness of farmers and their attitude toward new farm practices. Lindstrom (1958) found that "the adoption of practices must be preceded by the formation of favorable attitudes toward them."

Research by Hovland and others (Hovland & Janis, 1959) on persuasion and attitude change indicates that the higher the credibility of the communicator, the more likely the opinion change. The indication that ideas are accepted more readily from high prestige persons than from low, supports the proposition for, by the definition of prestige, the receiver of the idea must perceive the sender as having a structure compatible with his own, and this compatability is projected onto the innovation. Research from the same area indicates that a strong threat may be less effective than a mild threat in inducing opinion change. A mild threat is an indication of high compatibility with the individual's structure and thus positively related to innovativeness.

Stogdill (1959) expresses the notion of this proposition at the group level when he says:

Both individuals and subgroups are likely to oppose a change which challenges the legitimacy of their roles, dislodges them from their positions, reduces the value of the outcomes they can experience in the group, imposes unnecessary hardships, or places them at a comparative disadvantage to other individuals or subgroups [p. 288].

The Junior High School study of Willower and Jones (Willower, 1962) indicated a resistance by staff members to changes which would benefit them at the expense of other members. Carter and Williams (1957) report that:

To a great extent the receptiveness of a firm to technical change must depend on the understanding and appreciation of science and technology to be found in the various ranks of management; and this, in turn, depends on the training and experience of the individual managers [p. 66].

The classic research of Coch and French (1948) indicated that by involving staff in development of change plans, resistance to that change could be greatly reduced. These results may be interpreted as further confirmation of the proposition.

Miles (in Meierhenry, 1964) summarizes these ideas in the following statement:

However, other things being equal, innovations which are perceived as threats to existing practice rather than mere additions to it are less likely of acceptance. More generally, innovations which can be added to an existing program without seriously disturbing other parts of it are likely to be adopted [p. 208].

The effect of cultural norms on the acceptance of innovations is illustrated by a study conducted by Pedersen (1951). He found that a Danish community accepted new agricultural practices more readily than a Polish community. Rogers (1962) suggests that the reason for this difference in readiness to accept innovations lies in the fact that:

The Polish farmers came from a background of subsistence farming; the Danish farmers were accustomed to producing for a world food market. The cultural values of the Danish farmers facilitated the adoption of new ideas, whereas the norms of the Polish community perpetuated the *status quo* [p. 59].

One might also suggest that Weber's (1958) thesis on the relationship between the Spirit of Capitalism and the Protestant Ethic was an example of the adoption of an innovation (Capitalism) because of the compatibility of it with the structure (Protestant Ethic) of the culture.

Proposition 3. The greater the pervasiveness of the innovation within the system, the lower the system innovativeness.

Here "pervasiveness" is used in two ways. First, as the intensity of the effect on subsystems, and second, as the number of subsystems affected.

This proposition is clearly closely related to Proposition 2. However, it is possible to differentiate conceptually between a qualitative effect on system structure and a quantitative effect.

Proposition 3 suggests that an innovation which would alter an individual's habits, beliefs, and attitudes would be resisted more than an innovation which affected only his habits.

Menzel (1960) found that physicians were likely to adopt highly pervasive innovations. Lionberger (1960) summarizes additional evidence in the following way. "Generally speaking, the more complex a practice, the more change it requires in existing operations, the more slowly it will be adopted [p. 104]."

These first three propositions have dealt mainly with the input of information about innovations and their relationship to system innova-

tiveness. However, it is clear that input to one system is output from another. Therefore, innovativeness may also be conceived of as related to system output.

Proposition 4. The greater the output of the system to its environment, the lower its innovativeness.

The greater the output of a system to its environment, the stronger and more stable the system-environment interaction becomes, and thus, the greater the system's resistance to change. For, if a system has a great many "lines" of communication or interaction with its environment a change in it will upset the equilibrium or boundary-maintenance which has been established through each of these lines. If there are few of these connections, then changes will upset few system-environment equilibria.

The aforementioned study by Willower and Jones (Willower, 1962) tends to support this proposition. For instance, the resistance to changes, which would benefit some individuals at the expense of others, may be interpreted as indicating resistance related to the individual's affective output. That is, the more lines of affective output, the less innovative the individual.

Similarly, the indication by Carlson's (1965, p. 65) research that innovators tend to "know well fewer of their peers, (and to) be sought less often for advice" lends support to Proposition 4.

Gouldner (1959) discusses this notion within the concepts of interdependence and functional autonomy. He notes that "the problem of functional autonomy is of considerable significance for the analysis of tension within social systems, and thus for the analysis of social change [p. 255]." He further elaborates by saying:

That is, those parts with least functional autonomy, those which cannot survive separation from a social system, are more likely to be implicated in its conversion than those which can.

Contrariwise, those with most autonomy are most able to press for or to accept changes, when these are consistent with their own autonomy [p. 258].

Thus, sociometrically isolated individuals or isolated cultures are more susceptible to change than those which are integrated into their environments because the isolated systems have low interdependence.

The concept of "temporary systems," as developed by Miles (Miles [Ed.], 1964, pp. 454-456) has relevance here. He notes research which indicates the consequences of physical and social isolation of temporary systems. First, this isolation "tends to shear away the person's (or group's) preoccupation with, and allegiance to, "things as they are." Secondly, it reduces the role conflicts normally extant in permanent systems. Thirdly, it supplies a strong protective function. That is, "the penalties for making mistakes are reduced." Finally, the isolation of a

temporary system promotes a cohesiveness resulting from the individuals' sense of "being apart together."

If one accepts system isolation as lack of system output, then Miles' suggestion that temporary (isolated) systems change relatively easily would imply that permanent (non-isolated) systems are resistant to change. Thus, these ideas are also congruent with Proposition 4.

Proposition 5. The more the system perceives itself as compatible with its environment, the lower its innovativeness.

A system perceives itself as compatible with its environment when its needs are satisfied through inputs from the environment which are seen as resulting from its outputs to the environment. Therefore, compatibility of system and environment is a state toward which the system strives. Boulding (1956) suggests that the behaviour of a system can be explained in terms of attempts to restore these preferred states when they are disturbed by changes in the environment.

It follows, therefore, that if a system perceives itself as fairly compatible with its environment, there will be little tension or stress inducing change.⁴ On the other hand, if the system and its environment are not completely compatible there will be pressure for change. Furthermore, the feedback mechanisms will bring about change directed toward increasing compatibility.

Rogers (1962) discusses research from several traditions which suggests that crisis conditions in the environment affect system innovativeness. That is, crisis conditions create changes in the environment which upset system-environment compatibility. The most recent and notable crisis to affect North American education was the launching of the Soviet sputnik. Brickell (1961) says that the rate of instructional innovation in the schools of New York State more than doubled in the 15 months following this event.

The generally inconclusive results of research on the relationship between the adoption of farm practices and farmers' values and attitudes suggests further support for the proposition. For instance, Lionberger (1960, pp. 93-94) discusses research which shows that innovativeness of farmers is related to the kinds of models with which they identify. Thus it is the system's perception of its compatibility with the environment which determines innovativeness. One would not expect particular value orientations or attitudes to predict innovativeness. Rather, the perception of the compatibility of these attitudes with those of the environment would determine whether or not the system would adopt the innovation.

⁴ Contrary to this, McClelland *et. al.*, hypothesized that small displacements from the "adaptation level" would be favorably received by organizations (*The achievement motive*. New York: Appleton-Century-Crofts, 1953). Similarly, the work of Bexton, Heron, and Scott indicates that total compatibility of the individual with his environment may produce a state of arousal (Bexton, W. H., Heron, W., and Scott, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology*, 1954, Vol. VIII, 70-76).

Mansfield (1961) found that non-competitive firms were slower to innovate than competitive ones. Carlson (1965, pp. 6-7) calls these non-competitive organizations "domesticated." If one interprets lack of competition as compatibility with the environment, then this notion confirms the proposition.

March and Simon (1959) explain these phenomena in the following way:

Individuals and organizations give preferred treatment to alternatives that represent continuation of programs over those that represent change . . . persistence comes about primarily because the individual or organization does not search for or consider alternatives to the present course of action unless that present course is in some sense "unsatisfactory" [p. 173].

Proposition 6. The greater the pervasiveness of the output of a system, the lower its innovativeness.

Again, "pervasiveness" is used as both the intensity of effect on suprasystems, and as the number of suprasystems affected. This meaning of pervasiveness is very similar to that used by Etzioni (1961). He says, "the range of pervasiveness is determined by the number of activities in or outside the organization for which the organization sets norms [p. 163]."

Rogers (1962, p. 177) notes several research projects which indicate that more specialized systems tend to be more innovative. It would seem reasonable to suggest that specialization of a system could be considered an indicator of low pervasiveness of output, and thus this research tends to support Proposition 6.

The propositions which have thus far been presented have considered input and output of the system separately. It is obvious, however, that for any open system, the two are interdependent. The concept of feedback connects system input to output, and suggests a final proposition.

Proposition 7. The greater the ratio of positive to negative feedback of the system, the higher the system innovativeness.

Although systems tend toward steady states, it is clear that systems do change. The changes must, therefore, be a result of variation of input. If a system is in *stationary equilibrium* the negative feedback is sufficiently great to overcome any positive feedback, and the system will return to its original state despite a change in input. If, however, the positive feedback is greater than the negative feedback, the system will not return to its original state after a change of input, but will achieve a different equilibrium. A system continuously undergoing this type of process is said to be in *dynamic equilibrium* (Chin, 1961). In the terminology used here, a system in dynamic equilibrium would be more innovative than a system in stationary equilibrium.

Although positive and negative feedback have not been researched *per se*, the wealth of studies at Columbia University has shown that the adaptability (a synonym for innovativeness) of school systems is very largely a function of the characteristics of the population of the community. Mort (1964) summarizes this research as follows:

Communities vary in the degree to which they take on new practices. Indications are that this is a community characteristic. A community that is slow to adopt one innovation tends to be slow to adopt others. A pioneer in one area tends to be a pioneer in other areas.

Explanation of the difference in educational adaptability of communities can be found in no small degree in the character of the population, particularly in the level of the public's understanding of what schools can do, and citizens feeling of need for education for their children. This appears to set the posture of the community toward financial support, and toward what teachers are permitted to do—and tends to shape the staff by influencing personnel selected and kept in the community [p. 326].

To the extent that community understanding, financial support and permissiveness can be considered as representative of the ratio of positive to negative feedback, this research tends to support the proposition.

Predominance of negative feedback would, according to the proposition, result in low innovativeness. This phenomenon was noted in a study reported by Eichholz and Rogers (1964). They found that ninety-three per cent of the teachers in their sample had rejected or discontinued the use of electromechanical innovations through experience. That is, the experience of the teacher with the innovation had resulted in negative feedback sufficient to mitigate against its future use.

Characteristics of Innovation

The theory of innovativeness presented here would appear to suffer from one serious omission. At no time has there been direct consideration given to the characteristics of the innovation. This omission was not unintentional. Characteristics of innovations which have been shown to be related to innovativeness appear to be closely associated with the propositions presented.

Rogers (1962, pp. 124-134) suggests that the (a) relative advantage, (b) compatibility, (c) complexity, (d) divisibility, and (e) communicability of an innovation affect its adoption. These characteristics would seem to be implied by, if not included in, the stated propositions. The communicability of the innovation is clearly implied in Proposition 1. The complexity and divisibility of the innovation are obviously related to the extent of change in the system required for adoption, and therefore, implied by Proposition 3. Finally, the compatibility and relative advantage of the innovation represent the compatibility of the innovation with the structure of the system, Proposition 2.

Conclusions

It is clear that the propositions stated here have been discussed as relatively discrete relationships. However, they are obviously statements of interdependent concepts and relationships. Of particular interest is the paradoxical symmetry of Propositions 1 to 3, and 4 to 6. The second triad hypothesizes relationships between certain output concepts and innovativeness which are to a certain extent converse to those implied in the first triad. For instance, Proposition 1 suggests that innovativeness rises when innovation information input rises, whereas Proposition 4 states that the greater the output from a system to its environment the lower its innovativeness. And yet the assumption of feedback structures means that there is a necessary and direct connection between system output and system input. If this is the case, then the relationships proposed in the first triad will tend to negate the relationships in the second triad, and the system will remain static. The result is basic system resistance to innovation.

Nevertheless, systems do change, and they do adopt innovations. Therefore, the relationships hypothesized must have differential magnitudes which allow innovativeness. It is suggested that the theory presented here is capable of accounting for the "orderly reciprocity of persistence and change (Williams, 1966, p. 15)."

Further, it is suggested that this theory is sufficiently general in nature to encompass the notions and theories of change in the disciplines of psychology, social psychology, sociology, and anthropology. For instance, at the individual level, Festinger's (1957) theory of cognitive dissonance is closely related to Proposition 1. At the group level, the principles of change suggested by Cartwright (1951) may be interpreted in terms of these propositions.

Theories of acculturation or cultural innovativeness may also be accounted for by the present theory. For instance, Toynbee's (1933) concept of "challenge and response" in societal development is clearly related to Proposition 5. Barnett (1953) states these hypotheses:

... an individual will not accept a novelty unless in his opinion it satisfies a want better than some existing means at his disposal.

* * *

... people develop tastes and preferences under the influence of particular experiences, and these orientations are significant for the acceptance or rejection of new ideas [pp. 378-379].

These appear to be specific cases of Proposition 2.

Similarly, Moore (1959) notes a number of acculturative situations of readiness and capability for novelty (innovativeness in the terms used here).

1. The greater the number of lines of contact with an external culture, the greater is the force of the new influence.
2. Acceptance of novelty is much greater and easier if contact with the new element is continuous and, in principle, permanent.
3. Under equivalent or analogous conditions simple elements are more readily transferable than complex ones and simple cultural forms are more transferable than complex cultural meanings.
4. Acceptance is greater and easier if congruent with the sense of advantage of those with recognized vested interests.
5. The higher the prestige level of the innovators or donors, the faster the rate of acceptance.
6. The rate and areas of acceptance of novelty are a function of the intensity and the directions of prevailing expectations of change.
7. Acceptance varies with the tension or gratification level which characterizes relationships generally between the innovator-donor and the acceptor cultures [pp. 179-198].

It is clear that one could also place these seven statements in the framework of propositions on the general theory developed here, as one could many of the theories of learning, role behaviour, leadership, and power and authority.

In addition to its generality in terms of previous theories, this theory appears to present a framework upon which many of the research findings on the adoption and diffusion of innovations may be placed. There is no doubt, however, that the concepts and relationships which have been presented here require confirmation through further empirical investigation. It is the writer's contention that these concepts and propositions can be "operationalized" and studied in diverse situations. For, as Getzels (1958) suggests, "in the long run, the best test of the usefulness of a theoretical formulation is whether or not it generates research and inquiry [p. 150].

References

- Andrews, J. H. M., and Greenfield, T. B. Organizational themes relevant to change in schools. *Ontario Journal of Educational Research*, Vol. IX, No. 2, 81-99.
- Barnett, H. G. *Innovation: The basis of cultural change*. New York: McGraw-Hill, 1965.
- Bennis, W. G. *Changing organizations*. New York: McGraw-Hill, 1966.
- Boulding, K. E. General systems theory—the skeleton of science. In L. Von Bertalanffy, and A. Rapoport (Eds.), *General Systems*. Yearbook for the Society for the Advancement of General Systems Theory, 1956, Vol. I, 18.
- Brickell, H. M. *Commissioner's 1961 catalog of educational change*. Albany, New York: State Education Department, 1961. P. 27.
- Carlson, R. O. *Executive succession and organizational change*. Chicago: Midwest Administration Center, The University of Chicago, 1962. P. 69.
- Carlson, R. O. *Adoption of educational innovations*. Eugene, Oregon: Center for the Advanced Study of Educational Administration, The University of Oregon, 1965.
- Carter, C. F., and Williams, B. R. *Industry and technical progress*. London: Oxford University Press, 1957. Pp. 178-180.

- Cartwright, D. Achieving change in people: Applications of group dynamics theory. *Human Relations*, 1951, Vol. IV, 388-391.
- Chin, R. The utility of system models and developmental models for practitioners. In W. G. Bennis, K. D. Benne, and R. Chin (Eds.), *The planning of change*. New York: Holt, Rinehart & Winston, 1961. Pp. 201-214.
- Coch, L., and French, J. R. F., Jr. Overcoming resistance to change. *Human Relations*, 1948, Vol. I, 512-532.
- Coleman, J., Katz, E., and Menzel, H. The diffusion of an innovation among physicians, *Sociometry*, 1957, Vol. XX, 253-270.
- Eicholz, G., and Rogers, E. M. Resistance to the adoption of audio-visual aids by elementary school teachers: Contrasts and similarities to agricultural innovations. In M. B. Miles (Ed.), *Innovation in education*. New York: Bureau of Publications, Teachers' College, Columbia University, 1964. Pp. 299-316.
- Etzioni, A. *A comparative analysis of complex organizations*. New York: The Free Press of Glencoe, 1961. P. 163.
- Festinger, L. *A theory of cognitive dissonance*. Evanston, Illinois: Row, Peterson, 1957.
- Fliegel, F. C. A multiple correlation analysis of factors associated with adoption of farm practices. *Rural Sociology*, 1956, Vol. XXI, 284-292.
- Getzels, J. W. Administration as a social process. In A. W. Halpin (Ed.), *Administrative theory in education*. Chicago: The Midwest Administration Center, The University of Chicago, 1958. P. 150.
- Gouldner, A. W. Reciprocity and autonomy in functional theory. In L. Gross (Ed.), *Symposium on sociological theory*. New York: Harper & Row, 1959. P. 255.
- Hall, A. D., and Fagen, R. E. Definition of system. In L. Von Bertalanffy, and A. Rapoport (Eds.), *General Systems*. Yearbook for the Society for the Advancement of General Systems Theory, 1956, Vol I, 18.
- Hovland, C. I., and Janis, F. I. (Eds.), *Personality and persuability*. New Haven, Connecticut: Yale University Press, 1959.
- Lazarus, R. S. *Personality and adjustment*. Englewood Cliffs, N.J.: Prentice-Hall, 1963. P. 67.
- Lindstrom, D. E. Diffusion of agricultural and home economics practices in a Japanese rural community. *Rural Sociology*, 1958, Vol. XXIII, 174.
- Linton, H., and (Sofer) Graham, E. Personality correlates of persuability. In C. I. Hovland, and F. I. Janis (Eds.), *Personality and persuability*. New Haven, Connecticut: Yale University Press, 1959. P. 70.
- Lionberger, H. *Adoption of new ideas and practices*. Ames, Iowa: The Iowa State University Press, 1960.
- Mansfield, E. Technical change and the rate of imitation. *Econometrica*, 1961, Vol. XXIX, 741-766.
- March, J. G., and Simon, H. A. *Organizations*. New York: Wiley, 1959. P. 173.
- Menzel, H. Innovation, integration and marginality: A survey of physicians. *American Sociological Review*, 1960, Vol. XXV, 704-713.
- Miles, M. B. Educational innovations: Some generalizations. In W. C. Meierhenry (Ed.), *Media and educational innovation*. Lincoln, Nebraska: University of Nebraska Press, 1964. P. 208.
- Miles, M. B. (Ed.), *Innovation in education*. New York: Bureau of Publications, Teachers' College, Columbia University, 1964. Pp. 425-436.
- Miles, M. B. Planned change and organizational health: Figure and ground. In R. O. Carlson et. al., *Change processes in the public schools*. Eugene, Oregon: Center for the Advanced Study of Educational Administration, University of Oregon, 1965. Pp. 11-34.

H. David Hemphill

- Miller, J. G. Living systems: Cross-level hypotheses. *Behavioral Science*, 1965, (October), Vol. X, 380-411.
- Moore, W. F. *Industrialization and labor: Social aspects of economic development*. Ithica, New York: Cornell University Press, 1959. P. 179.
- Mort, P. R. Studies in educational innovation from the institute of administrative research: An overview. In M. B. Miles (Ed.), *Innovation in education*. New York: Bureau of Publications, Teachers' College, Columbia University, 1964. P. 326.
- Pedersen, H. A. Cultural differences in the acceptance of recommended practices. *Rural Sociology*, 1951, Vol. XVI, 37-49.
- Riesman, D., et. al. *The lonely crowd*. New Haven, Connecticut: Yale University Press, 1961.
- Rogers, E. M. *Diffusion of innovations*. New York: The Free Press of Glencoe, 1962.
- Stogdill, R. M. *Individual behavior and group achievement*. New York: Oxford University Press, 1959. P. 288.
- Toynbee, A. J. *The study of history*. London: Oxford University Press, 1933. 12 Vols.
- Von Bertalanffy, L., and Rapoport, A. (Eds.), *General Systems*. Yearbook for the Society for the Advancement of General Systems Theory, 1956, Vol. I.
- Weber, M. The protestant ethic and the spirit of capitalism, (trans. Talcott Parsons). New York: Scribner, 1958.
- Williams, T. R. The study of change as a concept in cultural anthropology. *Theory into Practice*, 1966, Vol. V, 15.
- Willower, D. J. Barriers to change in educational organizations. *Theory into Practice*, 1962, (December), Vol. II, 258.

This article summarizes evidence of the construct validity of several "culture-reduced" measures of intellectual ability for natives. Three characteristics of such tests suggest further research. An additional conclusion is that large proportions of Canadian native pupils of early school age have the general intellectual ability which seems necessary to participate fully in the larger Canadian community.

R. S. MacARTHUR
The University of Alberta

Assessing Intellectual Potential of Native Canadian Pupils: A Summary

Current emphasis on the development of human resources has underlined the need for valid estimates of the intellectual potential of individuals from a variety of backgrounds to aid in adapting teaching or training to that potential. Now that some long-term predictive validity data are available, this article summarizes evidence from a series of studies of the construct validity of several "culture-reduced" measures of general intellectual ability for native pupils of the Canadian West and North.

Rationale

A suggested rationale to guide such studies has been outlined elsewhere by MacArthur (1968) and by West and MacArthur (1964). In brief, it conceptualizes abilities as organized in a hierarchy from relatively specific abilities at the bottom to general intellectual ability at the top, resembling the models of Vernon (1965) and of Gagné (1962). The development of these abilities takes place through a sort of cumulative transfer as innate predispositions interact with environmental conditions, along lines suggested by Hebb (1949) and by Piaget (1964). Since environmental conditions may differ considerably from one culture to another, so may the patterning and nature of abilities at all levels of the hierarchy, as suggested by Ferguson (1954) and by Biesheuval (1962).

Abilities high in the hierarchy are less affected by particular environmental experiences, and hence measures of general intellectual ability with minimum cultural bias should provide least bad estimates of intellectual potential for individuals from other cultures.

This line of reasoning has led to eight criteria suggested for the construct validity of a measure of general intellectual potential for pupils who are not from middle-class urban European-American backgrounds.

1. It should largely sample the broad factor of general intellectual ability running through a variety of European-American kinds of intellectual tasks, including reading, linguistic, and arithmetic achievement.

2. It should show less difference between cultures in cross-cultural administration than do alternative measures of intelligence. This is the concept of a "culture-reduced" test (as contrasted with the impractical concepts of "culture-free" or "culture-fair"), in the sense of one that minimizes but does not eradicate the effects of bias in terms of a particular cultural environment.

3. It should show moderate relationship with concurrent school achievement or trade efficiency.

4. It should show long-term validity as a predictor of success in intellectual tasks when appropriate adaptive intervening treatments have been employed.

5. It should be reliable.

6. It should show stability under changed environmental conditions, relative to that of alternative measures of intelligence.

7. It should minimize effects of test sophistication, providing plenty of appropriate practice experience, having directions depending little if at all upon language, and be unspeeded.

8. It should be administratively practical in that it has face validity and intrinsic interest, it is economical, it can be administered to groups by competent teachers, and it has appropriate norms or is amenable to the production of such norms.

Data Collection

Samples

Data are here summarized for three groups of samples. Group I consisted of 126 Metis pupils in Grades 1 to 8 at Faust, Alberta, with replication on 155 Metis and Indian pupils in Grades 1 to 9 at Fort Simpson, N.W.T. These pupils were sub-divided into eight samples for analysis by grade-levels, and are described in MacArthur (1962). Group II, the prediction samples, consisted of the 45 Metis pupils who were in Grades 1 to 3 in the above Faust samples and were attending school at Faust four years later, and of 97 Eskimo pupils in two groups ages 6 to 9 and 9½ to 12 years, at Inuvik and Tuktoyaktuk, for whom three-year predic-

tion data were available. These samples are described in Rattan and MacArthur (1968). Group III, the norming samples, consisted of 792 Indian-Metis, 510 Eskimo, and 709 Whites who were representative of pupils attending Grades 1 to 8 of the schools of the Mackenzie District, N.W.T. The samples are described in MacArthur (1965).

Measures

Among the group ability tests administered to the samples of Group I were those listed in Table 1, the first five of which were hypothesized to be somewhat culture-reduced measures of general intellectual ability, and the others of which include conventional measures of scholastic aptitude and school achievement. The actual tests used, with levels considered appropriate for the respective grade-levels, were: Progressive Matrices (Coloured and Standard), Safran Culture-Reduced Intelligence Test Scale 1, IPAT Cattell test of g Scale 2, Lorge-Thorndike Non-Verbal Intelligence Test (four levels with three sub-tests and a total score for each level), California Short-Form Test of Mental Maturity (three levels with six sub-tests and a Total score for each level), Detroit Beginning First-Grade Intelligence Test, Otis Quick-Scoring Mental Ability Tests (Alpha and Beta), and California Achievement Battery (four levels with three sub-tests and a Total score for each level). A selection from these tests was administered to the prediction samples of Group II and the norming samples of Group III.

Results and Discussion

For each sample of Group I separately, after normalizing most scores, the intercorrelations of the variables listed in Table 1 were principal-factor analyzed, with unities inserted in the diagonal cells and common factors whose corresponding eigenvalues were greater than one considered significant. Where no entry for a test occurs, that test was not administered to the particular sample. The unrotated first principal factor loadings shown in Table 1 may be considered to indicate the degree to which a test is measuring, for a sample, the most important single intellectual ability factor running through these cognitive tasks which include conventional verbal intelligence tests and school achievement tests. The communalities shown in Table 2 may be considered to indicate the degree to which a test is measuring, for a sample, all of the various common factors running through the battery of ability and school achievement tests. In later studies similar results were obtained for several Eskimo samples. Tables 1 and 2 taken together indicate the degree to which the various measures are meeting Criterion 1 for culture-reduced measures of general intellectual ability. (Rotated factor patterns are reported in MacArthur (1962). Other factor analyses were made using sub-tests of the Lorge-Thorndike and California Mental Maturity Tests.)

TABLE 1
UNROTATED FIRST PRINCIPAL FACTORS
Metis and Indian Samples

Sample	Gr. 7 & 8		Gr. 5 & 6		Gr. 2 & 3		Gr. 1	
	Faust	Simpson	Faust	Simpson	Faust	Simpson	Faust	Simpson
N	23	32	29	58	42	46	32	19
Mean Age	14-4	13-10	12-10	13-11	9-4	10-2	7-9	7-6
Prog. Matrices	.76	.59	.88	.51	.82	.46	.64	.76
SCRIT	.55	.55	.81	.56	.79	.67	.73	.71
IPAT Cattell	.75	.55	.78	.69				
Lorge-Th. NV Tot.	.77	.86	.89	.82	.76	.73	.75	.73
CTMM Non-Lang.	.60	.32	.34	.57	.75	.67	.36	.88
CTMM Lang.	.86	.79	.76	.71	.83	.62	.58	.92
Detroit Begin.					.63		.49	.84
Otis Beta	.79	.80	.63	.88				
Otis Alpha NV						.71		
Otis Alpha Vbl.						.73		
TOGA Total	.70		.82		.47		.68	
Calif. Ach. Rdg.	.82	.87	.60	.86	.90	.74	.69	.94
Calif. Ach. Arith.	.80	.83	.62	.90	.80	.72	.81	.91
Calif. Ach. Lang.	.82	.87	.07	.81	.86	.78	.75	.68
Prop. Total								
Variance	.57	.53	.49	.55	.59	.48	.43	.67

TABLE 2
COMMUNALITIES
Metis and Indian Samples

Sample	Gr. 7 & 8		Gr. 5 & 6		Gr. 2 & 3		Gr. 1	
	Faust	Simpson	Faust	Simpson	Faust	Simpson	Faust	Simpson
No. Common Factors	3	3	4	3	3	3	3	2
Prog. Matrices	.72	.64	.80	.66	.89	.75	.51	.71
SCRIT	.91	.69	.79	.67	.80	.67	.57	.90
IPAT Cattell	.71	.77	.78	.66				
Lorge-Th. NV Tot.	.74	.77	.83	.82	.60	.65	.57	.55
CTMM Non-Lang.	.83	.94	.91	.87	.67	.71	.94	.76
CTMM Lang.	.86	.87	.68	.73	.71	.74	.63	.88
Detroit Begin.					.77		.81	.77
Otis Beta	.87	.87	.82	.84				
Otis Alpha NV						.59		
Otis Alpha Vbl.						.67		
TOGA Total	.73		.75		.87		.47	
Calif. Ach. Rdg.	.80	.89	.95	.87	.90	.75	.77	.88
Calif. Ach. Arith.	.79	.70	.75	.84	.87	.61	.73	.92
Calif. Ach. Lang.	.84	.77	.91	.75	.87	.85	.77	.63
Prop. Total								
Variance	.80	.79	.82	.77	.79	.70	.68	.78

For the tests for which entries appear in Table 3, scores were obtained on samples totalling 780 pupils representative of the Grade 7 pupils in Edmonton, and the Grades 6, 3, and 1 pupils of Calgary; these were converted to T-scores based on the respective White samples, so that the means for each White urban sample on each test all were 50 with standard deviation 10. Scores for the eight Indian-Metis samples of Group I were then expressed in terms of the most nearly appropriate White urban T-scores, with means for the natives as shown in Table 3. This table, the data of which are extracted from West and MacArthur (1962), thus shows the degree to which the various measures are meeting Criterion 2 (the tests with highest means in any column have the least cultural bias). It provides an empirical definition of the concept "culture-reduced", for these Metis and Indian samples.

TABLE 3
MEAN T-SCORES FOR METIS AND INDIAN SAMPLES
BASED ON EDMONTON OR CALGARY SAMPLES
(White urban means are 50, S.D. of 10, in each instance)

Sample	Gr. 7 & 8 in Edmonton T		Gr. 5 & 6 in Calgary T		Gr. 2 & 3 in Calgary T		Gr. 1 in Calgary T	
	Faust	Simpson	Faust	Simpson	Faust	Simpson	Faust	Simpson
1. Prog. Matrices	41	45*	37*	40*	35*	36*	41*	41*
2. SCRIT			36*	39*	38*	40*	49*	41*
3. IPAT Cattell	41	41*	38*	39*				
4. Lorge-Th. NV Tot.	42*	45*	31	27	33	24	39*	35*
5. Lorge-Th. NV Pt. 3	43*	46*	34	29	45*	37*	43*	40*
6. Lorge-Th. NV Pt. 2	41	45*	34	29	41*	40*	45*	39*
7. Lorge-Th. NV Pt. 1	47*	45*	37*	34*	29	20	38	32
8. CTMM Non-Lang.	43*	41*			41*	42*		
9. CTMM Spatial	46*	45*			48*	53*		
10. CTMM Logical	38	35			35*	30*		
11. CTMM Numerical	37	36			32	26*		
12. CTMM Verbal	36	33			28	17		
13. CTMM Lang.	33	30			26	17		
14. CTMM Total	36	31			29	21		
15. Detroit Begin.							35	27
	*Sig. Diff. from CTMM Total at .01 level		*Sig. Diff. from L-Th NV Total at .01 level		*Sig. Diff. from CTMM Total at .01 level		*Sig. Diff. from Detroit Begin. at .01 level	

Using median values over the eight samples of Group I, Table 4, which represents the core of this article, presents a summary evaluation of nine tests which were hypothesized to be culture-reduced measures of general intellectual ability in terms of: Criterion 1—(a) unrotated first principal factor loadings, and (b) communalities; Criterion 2 — mean T-scores based on White urban standards; Criterion 3 — correlation with concurrent school achievement as measured by California Achievement Tests Total. The crudity of such overall summaries using median values,

not always over the same numbers of samples, is emphasized; for detail the reader is referred to previous tables and to references already given. In general, however, it is seen that tests least badly meeting the first three criteria set up for measures of intellectual potential with minimum

TABLE 4
SUMMARY EVALUATION OF NINE CULTURE-REDUCED TESTS
IN TERMS OF FOUR CRITERIA

Criterion	1		2		3		4		
	Mdn. 1st Unrotated Factor	Mdn. Commun- ality	Mdn. of Mean T-scores	Mdn. r Concur- rent Schl. Ach.	45 Metis over 4 yrs.	62 Esk. over 3 yrs.	37 Esk. over 3 yrs.	Long-term r with Schl. Ach.	
1. Prog. Matrices*	.70	.72	41	.53	.57	.43	.60		
2. SCRIT*	.69	.74	40	.49	.62				
3. IPAT Cattell*	.72	.74	40	.43					
4. Lorge-Th. NV Tot.	.77	.70	34	.64					
5. Lorge-Th. NV Pt. 3	.65	.81	42	.51					
6. Lorge-Th. NV Pt. 2	.71	.74	41	.57					
7. Lorge-Th. NV Pt. 1	.60	.72	36	.48					
8. CTMM Non-Lang.	.59	.85	42	.41	.50				
9. CTMM Spatial	.43	.78	47	.29	.37				
10. CTMM Logical	.63	.71	35	.59	.59				
11. CTMM Numerical	.66	.80	34	.58	.52				
12. CTMM Verbal	.72	.82	31	.73	.54				
13. CTMM Lang.	.78	.73	28	.73	.68				
14. CTMM Total			30	.72	.64				
15. Detroit Begin.	.63	.79	31	.44					
16. Otis Beta	.79	.85		.76					

*Tests least badly meeting these criteria for construct validity.

cultural bias are Progressive Matrices, Safran Culture-Reduced Intelligence Test, Cattell test of g Scale 2, and some sub-tests of the Lorge-Thorndike Non-Verbal Intelligence Tests. The Non-Language and Spatial sub-tests of the California Test of Mental Maturity, though showing less cultural bias than other measures in the battery, are not as highly loaded on the first principal factor and do not relate as highly to school achievement. As might be expected, the conventional tests 10 to 16 of Table 4 are somewhat more highly related to school achievement, but show much more cultural bias against the Metis and Indian pupils.

Table 4 also summarizes, for some of the measures, data from Rattan and MacArthur (1968) with respect to Criterion 4 — long-term predictive validity for samples of Group II. For the 45 Metis in the prediction sample, Coloured Progressive Matrices and SCRIT did not differ significantly from any of the more conventional tests 10 to 14 in predicting California Achievement Total over a four-year period.

Reliabilities of the various tests were not studied directly for these particular samples. However, communalities provide minimum estimates

of reliabilities (reliabilities must be as high as, and may be much higher than, communalities). In general, all of the measures studied would appear to have satisfactory reliabilities. Stability coefficients for Progressive Matrices and SCRIT, for the Metis prediction sample, did not differ significantly over the four-year period from those of the conventional CTMM Total or Verbal, Numerical and Logical sub-tests. Directions for the two former tests rely heavily on gesture and demonstration, and their time limits were so set that all but the very slow pupils were able to finish; however, more experimental work with respect to Criterion 7 would be desirable.

All of the culture-reduced tests studied appear to be intrinsically interesting to the Metis, Indian, and Eskimo pupils, are relatively economical, and can be readily administered as group tests by competent teachers. Appropriate norms for such pupils often present a problem; on the basis of the above and similar evidence, the Canadian Department of Northern Affairs undertook the Mackenzie District Norming Project, the report of which by MacArthur (1965) presents age norms for Eskimo, Indian-Metis, and Whites separately, based on samples of Grades 1 to 8 pupils representative of the whole of the Mackenzie District of the North West Territories, for Coloured and Standard Progressive Matrices, Otis Alpha Non-Verbal, and Lorge-Thorndike Non-Verbal Levels 3 and 4.

Conclusions

1. The evidence summarized above would seem to suggest that, until much-needed research can improve on them, such culture-reduced measures of general intellectual ability as Progressive Matrices and SCRIT for all grades 1 to 8, and IPAT Cattell and Lorge-Thorndike Non-Verbal for the higher of these grades, should be included in testing programmes aimed at helping us understand and adapt instruction to the abilities and potential of individual native Canadian pupils. However, conditions and pupils are changing rapidly these days, and construct validities may change similarly. Continuing studies such as those summarized are needed.

2. One might ask—what is there about such tests as Progressive Matrices which helps them, at least for our samples, meet these criteria taken together less badly than do conventional group so-called intelligence tests? Three replies suggest themselves: (a) the items form something of an age-scale sampling stages in the development of human cognition, starting with perception-dominated items, and proceeding through reversible concrete operations, to propositional or formal operations; (b) they use as stimuli symbols, which, though dependent on learning, are likely to be learned in a variety of cultures; (c) arrangement of items in the test itself forms a crudely-programmed sample of learning-on-the-

spot. These suggest three directions in which continued basic research should help us with the practical problems of assessing the general intellectual potential of individual pupils or candidates from other cultures.

3. If the validity data presented here for Progressive Matrices are accepted, an additional conclusion from the Mackenzie District Norming Project is that large proportions of Canadian native pupils of early school age *have* the general intellectual ability to participate fully in the larger Canadian community. On this test, for seven-year-olds, for example, the 90th percentiles for Eskimo, Indian-Metis, and Whites were respectively 25, 25, 29; the 50th percentiles were 18, 16, 20.5; the 10th percentiles were 13, 9, 14. For nine-year-olds, the 90th percentiles for Eskimo, Indian-Metis, and Whites were respectively 31, 31, 33; the 50th percentiles were 22, 21, 26; the 10th percentiles were 14, 15, 18. Thus, of these young native pupils, many scored well above the average Whites. At ages 11 and 13 years few native pupils scored above average Whites, but many scored well above the 10th percentile for Whites.

These culture-reduced tests are still considered to be culturally biased when used in this context. Nevertheless these and other data from the references herein cited clearly indicate very considerable intellectual potential among the early-school-age native pupils of the Canadian West and North.

References

- Biesheuval, S. The detection and fostering of ability among underdeveloped peoples. *Yearbook of Education*, 1962, 337-353.
- Ferguson, G. A. On learning and human ability. *Canadian Journal of Psychology*, 1954, Vol. VIII, 95-112.
- Gagné, R. M. The acquisition of knowledge. *Psychological Review*, 1962, Vol. XXIX, 355-365.
- Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.
- Inhelder, B., and Piaget, J. *The early growth of logic in a child*. New York: Harper & Row, 1964.
- MacArthur, R. S. Assessing the intellectual ability of Indian and Metis pupils at Fort Simpson. Ottawa: Department of Northern Affairs and National Resources, 1962.
- MacArthur, R. S. Mackenzie norming project. Ottawa: Department of Northern Affairs and National Resources, 1965.
- MacArthur, R. S. Some differential abilities of Northern Canadian native youth. *International Journal of Psychology*, 1968, Vol. III.
- Rattan, M. S., and MacArthur, R. S. Longitudinal prediction of school achievement for Metis and Eskimo pupils. *Alberta Journal of Educational Research*, 1968, Vol. XIV, No. 1.
- Vernon, P. E. Ability factors and environmental influences. *American Psychologist*, 1965, Vol. XX, 723-733.
- West, L. W., and MacArthur, R. S. An evaluation of selected intelligence tests for two samples of Metis and Indian children. *Alberta Journal of Educational Research*, 1964, Vol. X, No. 1, 17-27.

Ziller's disguised measure of test-risk is compared with a reported measure of test-risk. Correlations are reported for the two measures, and the distributions are compared. The evidence presented supports the disguised measure as a valid assessment of risk taking.

MALCOLM J. SLAKTER and AUBREY H. RODEN
State University of New York at Buffalo

Comparison of a Disguised and Reported Measure of Risk Taking On Objective Examinations

In an objective achievement or aptitude examination, the examinee is confronted with a series of items, each of which contains a stem and a set of alternative responses. For each item, the examinee is directed to select the single best alternative for the given stem. In an attempt to diminish the effects of guessing upon test score, the examinee is often-times further instructed to omit those items to which he does not know the answer. The latter instruction comes in many forms and varieties, but all belong to the set of do-not-guess directions. Do-not-guess directions are generally accompanied by a "correction formula" or penalty for incorrect responses; i.e., test score equals the number of correct responses minus some fraction of the number of incorrect responses. Therefore, under do-not-guess directions, it is clear that whenever the examinee encounters an item in which he cannot select the "best" alternative with certainty, there exists a risk taking situation. If the decision is to omit the item, the use of the "correction formula" will not affect the test score. On the other hand, if the decision is to select one of the alternatives, the use of the "correction formula" will increase or decrease the test score, depending on whether or not the alternative selected is the best alternative.

There is ample evidence that when confronted with these risk taking situations, examinees differ in their risk taking propensities (Slakter,

1967a, 1967b; Stone, 1962; Swineford, 1938); i.e., on a given examination, examinees differ reliably in their tendency to omit items or to guess. It is reasonable to assume that this particular behaviour is an expression of a more general motivational pattern. Indeed, there is evidence that risk taking on objective examinations is related to behaviours such as vocational choice (Ziller, 1957a), dominance-submission (Votaw, 1936), and maladjustment (Sherriffs and Boomer, 1954).

In addition, there is evidence that the low risk taker on examinations tends to be penalized by his reluctance to respond to items that are not known with certainty (Sherriffs and Boomer, 1954; Slakter, 1967a; Votaw, 1936). These studies have demonstrated that when low risk takers are forced to respond to all items, their test score will generally increase even though the usual penalty for guessing is applied. Therefore, these findings provide evidence that the personality trait of risk taking confounds the achievement or aptitude being measured by the examination.

There have been several disguised measures of test-risk proposed in the literature (Danielson, 1956; Slakter, 1967a, 1967b; Swineford, 1938), but the most widely used to date has been that developed by Ziller (1957a). Ziller's index is a function of the number of incorrect and omitted items, and the number of alternatives per item. Specifically, each examinee receives a risk taking score R , where

$$R = \frac{\frac{n}{n-1} w}{\frac{n}{n-1} w + u},$$

and n is the number of alternatives for each item, w is the number of incorrect responses for the examinee, and u is the number of items unattempted by the examinee. For each subject, R is essentially an estimate of the ratio of items guessed to items not known. The range of R is from 0 to 1, with low values of R indicating low risk, and high values indicating high risk.

Another method of determining the extent of risk taking by an examinee is simply to ask him; i.e., a self-report. Naturally, this self-report suffers from the deficiencies of all self-reports: (a) the examinee must know the answer to the question posed, and (b) assuming he knows the answer, he must be willing to share the information with the person administering the self-report. Unfortunately, there is evidence that examinees are unable to clearly distinguish between those items on which they have complete, partial, or no information (Granich, 1931; West, 1923); i.e., examinees do not know precisely the extent of their test-risk. However, the lack of precise knowledge of degree of test-risk does

not necessarily imply that there is no knowledge. Clearly, the construct validity of the Ziller measure would be enhanced if there existed strong positive relations between the disguised Ziller test-risk score and the reported test-risk score. The main problem of interest in this paper, therefore, was to compare the Ziller index with a reported measure of test-risk. Specific areas of interest were (a) the correlation between the disguised and reported measures, and (b) a comparison of the distributions of the disguised and reported measures.

Method

The examination used in this study was the Concept Mastery Test (Terman, 1950) or CMT. The do-not-guess directions of the CMT instructs examinees to: "Omit those items that you could answer by pure guess, but answer all you *think* you know, even if you are not quite certain." The items in the first part of the CMT have two alternatives each, and are based upon vocabulary; the items in the second part of the CMT have three alternatives, and are based upon verbal reasoning. Since the two parts of the CMT have different numbers of alternatives, a slight modification is necessary in the calculation of R , the disguised measure. Therefore, in this study

$$R = \frac{2 w_1 + \frac{3}{2} w_2}{2 w_1 + \frac{3}{2} w_2 + u},$$

where w_1 is the number of incorrect responses on part 1, and w_2 is the number of incorrect responses on part 2.

Immediately following the administration and collection of the CMT, examinees were asked to complete a questionnaire which contained the following directions:

Please answer the following questions concerning the Concept Mastery Test that you have just taken. Your answers to these questions will have no effect whatsoever on your Concept Mastery Test score or on your grade in this course.

The questionnaire included the following items:

1. Consider those items on the test that you could answer by pure guess only. Approximately what percentage of these items did you answer?
2. Consider those items on the test that you thought you knew, even if you were not quite certain. Approximately what percentage of these items did you answer?

Responses to item 1 were considered to be the reported measure of risk taking (R'); responses to item 2 were regarded as a reported measure of risk taking with partial information (R'').

The study was replicated in five different classes in introductory educational psychology; three classes at the University of California at

Berkeley in the summer of 1965, and two undergraduate classes at the State University of New York at Buffalo during the spring semester of 1966. Females constituted approximately 75% of each class, with the Berkeley classes averaging about 27 years of age compared to about 21 years for the Buffalo groups.

Results

Table 1 presents the sample size, mean, and standard deviation in each class for the disguised (R) and reported (R') measures of risk taking, and the reported measure of risk taking with partial information (R''). The Berkeley classes are represented by C₁, C₂, C₃; the Buffalo classes are labelled B₁, B₂.

TABLE 1
SAMPLE SIZE, MEAN, AND STANDARD DEVIATION FOR DISGUISED (R)
AND REPORTED MEASURE (R') OF RISK TAKING, AND REPORTED MEASURE
OF RISK TAKING WITH PARTIAL INFORMATION (R'')

	C ^a ₁	C ₂	C ₃	B ₁	B ₂
<i>n</i>	26	40	17	51	32
X _R	.81	.75	.80	.72	.54
X _{R'}	.43	.40	.47	.28	.17
X _{R''}	.84	.88	.83	.71	.60
S _R	.21	.21	.20	.22	.22
S _{R'}	.45	.40	.37	.35	.23
S _{R''}	.27	.19	.29	.31	.30

^aC₁, C₂, C₃ represent Berkeley classes; B₁, B₂ represent Buffalo classes.

In each class, the mean disguised risk score was significantly different from the mean reported risk score (.05 level), with the data indicating that the mean of the disguised risk score was higher. In four of the five classes, the reported risk score was more variable than the disguised risk score.

The correlations among the disguised and reported measures of risk taking and the reported measure of risk taking with partial information, are provided in Table 2. Values with an asterisk are significantly greater than 0 at the .01 level for a one-tailed test.

TABLE 2
CORRELATIONS AMONG DISGUISED (R) AND REPORTED (R') MEASURES
OF RISK TAKING, AND REPORTED MEASURE OF RISK TAKING WITH
PARTIAL INFORMATION (R'')

	C ₁	C ₂	C ₃	B ₁	B ₂
γ _{RR'}	.67*	.70*	.67*	.57*	.44*
γ _{RR''}	.73*	.38*	.53	.22	.28
γ _{R'R''}	.46*	.32	.57*	.42*	.28

*Significantly greater than 0 at the .01 level for a one-tailed test.

Discussion

In Table 1 it is seen that in each of the five classes, the subjects displayed higher risk taking (R) than they reported (R'). There are at least three possible explanations for this discrepancy: (a) perhaps it is due to the failure of the subjects to recognize the extent of their risk, and hence the subsequent under-estimation of their self-report, (b) possibly the subjects recognized the extent of their risk, but for some reason (e.g., social desirability), they systematically reported lower risks than they took. Alternatively, it may be that (c) the subjects are accurately reporting their actual degree of risk taking, but for some reason(s) the disguised measure is overestimating the degree of risk. With respect to the last explanation, the assumption that each incorrect response indicates a pure guess provides two plausible reasons for the disguised measure being an overestimate of test-risk. The first reason is based on the effects of partial information, and the second on the effects of misinformation.

With partial information, the subject could not select the best alternative with certainty. Note that if the subject took a risk with partial information, an incorrect response would increase the disguised score (R), but the subject presumably would have taken account of the risk in the reported measure with partial information (R'') rather than in the reported measure of risk (R'). Hence, the mean of R would be inflated over that of R' . Indeed, it is interesting to note that in each class, the mean of the reported measure of risk with partial information was approximately the same as the mean disguised risk (see Table 1).

In the case of misinformation, the subject did not need to guess the alternative because he "knew" the answer—unfortunately for the subject, however, the response selected was not the one keyed by the test constructor. Therefore, while the incorrect response thus obtained would increase R , the response would not be considered a risk by the subject, and hence would not enter into either R' or R'' . Once again, therefore, the mean of R would tend to be higher than the mean of R' . Clearly, any of the three previous explanations could account for the observed mean discrepancy between R and R' . However, it would seem that the third explanation is the most plausible.

Table 2 provides evidence that the correlation between the disguised and reported measures of risk is substantial as well as statistically significant. The correlations ranged from .44 to .70, with a median value of .67; the common variance for the two measures therefore ranged from about .19 to .48, with a median value of .45. In examining the multiple correlation between R and both R' and R'' , it was found that no appreciable predictable variance in R was gained by the combining of R'' with R' . Hence it was decided that while risk taking with partial information may

shed some light on the mean discrepancy between R and R' , it provides essentially no further information on the variability of R .

In conclusion, evidence has been offered demonstrating that the correlation between the disguised and reported measure has a median value in the vicinity of .67. Because of this evidence, the construct validity of the disguised measure has been enhanced. While it is clear that the subjects attained higher mean disguised score than they reported, the effects of partial information and misinformation were offered as possible explanations for the observed mean discrepancy.

References

- Danielson, L. E. Gambling proneness: Its measurement and expression in examination situations. Unpublished doctoral dissertation, University of Michigan, 1956.
- Granich, L. A technique for experimenting on guessing in objective tests. *Journal of Educational Psychology*, 1931, Vol. XXII, 145-156.
- Sherriffs, A. C., and Boomer, D. S. Who is penalized by the penalty for guessing? *Journal of Educational Psychology*, 1954, Vol. XLV, 81-90.
- Slakter, M. J. The measurement and effect of risk taking on objective examinations. U.S. Department of Health, Education, and Welfare. Final Report: Project No. 5-8428, Contract No. 0E-6-10-239, 1967. (a)
- Slakter, M. J. Risk taking on objective examinations. *American Educational Research Journal*, 1967, Vol. IV, 31-43.
- Stone, L. A. Reliability of a utility of risk measure. *Psychological Reports*, 1962, Vol. X, 516.
- Swineford, F. The measurement of a personality trait. *Journal of Educational Psychology*, 1938, Vol. XXIX, 295-300.
- Terman, L. M. *Concept Mastery Test*. New York: Psychological Corp., 1950.
- Votaw, D. F. The effect of do-not-guess directions upon the validity of true-false or multiple-choice tests. *Journal of Educational Psychology*, 1936, Vol. XXVII, 698-703.
- West, P. V. A critical study of the right minus wrong method. *Journal of Educational Psychology*, 1923, Vol. VIII, 1-9.
- Ziller, R. C. A measure of the gambling response-set in objective tests. *Psychometrika*, 1957, Vol. XXII, 289-292. (a)
- Ziller, R. C. Vocational choice and utility for risk. *Journal of Counseling Psychology*, 1957, Vol. IV, 61-64.

Expectancy tables were constructed using the grade nine and twelve stanine grades on Provincial Examinations and the comparable rankings on the Scholastic Ability Test of students in a rural Alberta county. The use of the tables for guidance in a small high school is illustrated.

DALE G. ANDERSEN

Washington State University

and

HERBERT E. STILES

Didsbury, Alberta

The Use of Expectancy Tables in Educational Guidance

The rural high school often finds itself at a disadvantage in providing educational guidance services to its students because of the absence of a trained counsellor. Such was the case of the five high schools in a rural Alberta county during the past several years. However, the desire to offer the best that was possible under the circumstances led to a study involving the construction and use of several expectancy tables as guidance tools. This paper reports that study.

The expectancy table is a useful test interpretation instrument and educational guidance device even in the absence of a trained specialist. A Test Service Bulletin (1949) states that:

The expectancy table is not new; it has been known and used in the test field for more than a quarter of a century. . . . It is merely a grid containing a number of cells [p. 11].

Along the side are categories of standardized test results while along the top are the varying achievement divisions. For each person a tally mark is placed which shows, vertically, his score on a given standardized test, and horizontally, his subsequent achievement in some endeavour.

In the present study expectancy tables were constructed utilizing: (a) selected results of the June Examinations of the Examinations Branch, Department of Education, for grade nine students and (b) grade twelve status of those same students three years later.

Collection and Reporting of Data

The scores reported in stanines for each grade nine student in the County on the six June Examinations—Reading, Literature, Language, Social Studies, Mathematics, and Science—were collected for the years 1956 through 1961. The percentile rank for each student on the Scholastic Ability Test during the period from 1959 (when they were first administered) through 1961 was also gathered. These data were obtained from either the principals of the schools or from the Department of Education, Edmonton. During the period of the study many students entered County schools from outside the County. Data pertaining to these students were obtained by writing to the principal of the school where the student had written the final grade nine examinations. Few requests were ignored, but the information received was not always complete. Therefore the total number of subjects (N) reported in the tables in this paper varies somewhat.

The grade twelve status of subjects including the results of Grade Twelve Examinations was obtained from the principals of the high schools within the County. The period here was for 1959 through 1964, the period when the original grade nine students had advanced to the twelfth grade.

The first expectancy table constructed, Table 1 below, was for “Total of Stanines.” For the six year period of the study the stanines for each grade nine student on the six tests were added. This resulted in actual totals varying from 15 to 54. It was decided to group scores in intervals of five for ease in reporting, thus creating eight possible horizontal rows on the expectancy table.

TABLE 1
EXPECTANCY TABLE FOR TOTAL OF STANINES

Score (Total of Stanines)	N	Grade Twelve Status		
		Matriculation Pass (%)	Diploma Pass (%)	Neither (%)
50-54	19	100	0	0
45-49	57	77	18	5
40-44	99	44	38	18
35-39	121	23	47	30
30-34	178	8	34	58
25-29	125	3	19	78
20-24	74	0	4	96
15-19	14	0	0	100
Total N	687			

Grade twelve achievement entries were made on all expectancy tables according to three categories: (a) “Matriculation Pass” denoting the student attained scores of at least 50% on each Grade Twelve Examination and an average of at least 60% on all six Examinations, (b) “Diploma

Pass" showing the student attained at least 40% on Grade Twelve Examinations and had accumulated 100 high school credits, and (c) "Neither" which indicated the student did not meet the requirements of either of the first two categories.

After compiling raw totals for each cell on the table these were converted to percentages by row. It was felt this would facilitate interpretation and allow the data to be viewed in a more practical and usable fashion.

In reading Table 1 it is apparent that for students having a total stanine score between 56 and 54 inclusive, 100% of students in the past have attained matriculation. A further illustration is that of students with total scores of stanines between 20 and 24 inclusive in the past none have attained matriculation status, 4 in 100 have gotten a diploma, and 96 in 100 have neither matriculated nor attained a diploma pass.

Table 2 uses the stanine for an individual test, Language. This expectancy table was constructed as before except the row values range only from one to nine since only one test is involved.

TABLE 2
EXPECTANCY TABLE BASED ON LANGUAGE STANINE

Language Stanine Score	N	Grade Twelve Status		
		Matriculation Pass (%)	Diploma Pass (%)	Neither (%)
9	30	96	0	4
8	73	56	31	13
7	97	43	35	22
6	139	17	45	38
5	182	10	28	62
4	112	2	15	83
3	44	5	16	79
2	16	6	6	88
1	4	0	0	100
Total N	697			

The Mathematics stanine was used by itself for constructing Table 3. It would appear that this Examination score is more discriminating than the Language score since the delineations are more clear cut especially at the lower stanine levels.

The last of the expectancy tables (Table 4) that will be presented here involves fewer students. The Scholastic Ability Test (SAT) was administered only throughout the final three years of this study. The standardized test results utilized in this table claimed to reflect ability rather than achievement. An interesting observation in regard to this table is that of those who ranked in row nine of the SAT, the highest rank, only 80% matriculated.

TABLE 3
EXPECTANCY TABLE BASED ON MATHEMATICS STANINES

Mathematics Stanine Score	N	Grade Twelve Status		
		Matriculation Pass (%)	Diploma Pass (%)	Neither (%)
9	29	90	10	0
8	51	62	28	10
7	115	40	40	20
6	156	20	42	38
5	183	11	30	59
4	108	4	13	83
3	46	0	0	100
2	11	0	0	100
1	0	—	—	—
Total N	699			

TABLE 4
EXPECTANCY TABLE BASED ON SCHOLASTIC ABILITY TEST

SAT Score ^a	N	Grade Twelve Status		
		Matriculation Pass (%)	Diploma Pass (%)	Neither (%)
96-99	25	80	4	16
89-95	37	57	35	8
77-88	82	30	35	35
60-76	67	16	36	48
40-59	90	5	32	63
23-39	68	0	16	84
11-22	29	0	7	93
4-10	12	0	0	100
1-3	2	0	0	100
Total N	412			

^aThese are percentile rank scores and were arranged to be roughly equivalent to the stanine intervals used in Table 2 and Table 3.

Guidance Application

The intent of presenting the data shown here to students in the schools of the County is not to tell them what they must do nor to impose decisions on them as to what they should do. Guidance and counselling activities should always be pointed toward human self-determination and the acceptance by the individual of self-responsibility. Frequently, however, students, due to a lack of self-knowledge and inability to assess chances of future success, drop out of school or make choices without any awareness of the probabilities of success or failure involved. They do not realize even in a relative sense their potential to achieve. Some cases of this type would surely be avoided through appropriate presentation of facts.

It is not always the poor student, or the one lacking intelligence who is a dropout. According to French (1966) who conducted a study of dropouts in Ohio schools:

The greatest loss, . . . is to the individual who is restricted in development and whose contribution of talent to our way of life is thereby reduced. The restriction of self-realization that accompanies withdrawal from school before a student's capability for entering an appropriately high vocational activity is one of the most devastating aspects of the nation's dropout problems [p. 123].

Self-understanding is not the cure-all for the highly complex problem of dropouts, but it may be the answer for many students. It is especially important today because society is asking that students be retained in school for longer periods. The Alberta Legislature, 1966, has decreed that the compulsory school attendance age be raised from 15 to 16 years. This is going to create many problems for the student who is unable to conceptualize himself and his world realistically.

A study by Mueller and Rothney (1960) found that guidance conferences were necessary if superior students were to choose wisely when educational decisions must be made. Choices, after all, are made on the basis of the chooser's prediction of outcomes.

In using the expectancy tables as guidance tools one must be objective. Patterson (1962) states:

The results of the tests must be allowed to speak for themselves after adequate explanation of the meaning of the scores by the counselor. An illustration of what is meant by objective presentation of the results is the interpretation of a score on a college aptitude test. The statement, "Three out of four persons with scores like yours do not complete the first year of college," is objective. The statement, "Your score indicates that the chances are three out of four that you will not complete the first year of college," while apparently objective, is somewhat loaded. But the statement, "With a score such as yours should not attempt college," is definitely evaluative and judgmental and goes beyond the presentation of information to the client [p. 155].

Since the expectancy tables provided through the present study are from achievement and ability tests, they should be quite reliable, but the person making the interpretation must be objective when using them.

The completed tables may also be used in guidance when dealing with a class or with a group of students. A large copy of each table or an overhead projector transparency could be made. A discussion of each table may be given with explanations and results from previous classes. Students should be encouraged to discuss this information. This would facilitate understanding on the part of each individual as to what the facts are and more important may stimulate individuals to seek out more self-information and information about opportunities. This is especially important in the small secondary school where a guidance counsellor is not on staff.

For some time one of the co-authors of this paper, a high school principal in the County has used the tables presented here to offer as-

sistance to both parents and students. Two illustrations may help in conveying how the tables might be used.

Case One

A boy entering grade twelve with a stanine total of 25 from grade nine examinations was very frustrated. His achievement from the last year was in the low "C" category. He wished to leave school and take technical training in welding and go farming. The parents wanted the boy to complete grade twelve, go to university, and enter a faculty of law.

The mother came to school to talk to the principal. By using the expectancy chart (Table 1), it was possible to show that her son had 3 chances in 100 of matriculating and being able to enter university, and also that her son had 19 chances in 100 of receiving a diploma. The principal also explained that the boy's percentile rank of 48 meant only 37 chances in 100 of receiving a diploma.

The parents' decision after talking to their son, later, was that he could follow his own plan. The boy decided to stay in school and with the reduced parental pressure he had a much more successful year in school.

Case Two

This case involved a boy passing into grade ten in September with a stanine four in mathematics. His percentile rank was 66 on the SAT. The mother wanted the boy to take the academic mathematics.

Using the expectancy table for mathematics stanines she was shown that her son would have 4 chances out of 100 of succeeding in it. However, the mother felt that the general mathematics program was an inferior program and that her son had the ability to take the academic mathematics.

The principal accepted the mother's decision, but felt that the work would be very difficult for the boy as too much background was missing. The boy failed his mathematics in June.

Summary and Implications

This paper has reported the employment of expectancy tables in the educational guidance of rural youth in one county in Alberta. A description of the procedure in constructing these tables based on data from grades nine and twelve was offered and some guidance application reviewed. Two illustrative cases were sketched. The intention of the study was not to work out a device which would give a rigid rule for laying out a student's educational future. It was intended that the expectancy tables would give a realistic appraisal of the situation to the student, parent, or teacher. The greatest value of the tables, then, is as a guidance tool.

The expectancy tables could be refined so that the predictive value would show success of boys and girls separately. The tables could also be modified to be double-expectancy tables¹ combining total stanine scores from grade nine final examinations and the SAT percentiles.

It is hoped that the present paper has presented a procedure and some guidelines that will serve as models for similar projects in other secondary school districts in Alberta. The tables constructed proved useful in the district of their origin.

References

- French, J. L. High school dropouts of high ability. *The Vocational Guidance Quarterly*, 1966, Vol. XIV.
- Mueller, W. J., and Rothney, J. W. M. Comparison of selected descriptive and predictive statements of superior students, their parents and their teachers. *Personality and Guidance Journal*, 1960, Vol. XXXVIII, 621-625.
- Patterson, C. H. *Counseling and guidance in schools*. New York: Harper, 1962.
- Expectancy tables—a way of interpreting test validity. *Test Service Bulletin No. 38*. New York: Psychological Corporation, 1949.

¹ For an explanation of double-expectancy tables see: "Double-entry expectancy tables." *Test Service Bulletin No. 56*. New York: Psychological Corporation, 1966.

Two of the most basic theoretical constructs used by Piaget to explain the acquisition of conservation were identified, described, then tested for their ability to account for the child's performance on a test of conservation. The two constructs were the understanding of transformations and the understanding of states. It was hypothesized that Construct One was more important than Construct Two. This was not found.

DAIYO SAWADA and L. DOYAL NELSON
The University of Alberta

Transformations: Length Conservation

Since the invention by Piaget of using conservation as one important criterion for assessing the degree to which a child has acquired certain fundamental concepts of science and mathematics, much research has been undertaken for the purposes of: (a) replicating Piaget's investigations, (b) exploring the efficacy of inducing conservation by various training methods, (c) attempting to ascertain the effects of other variables such as age, IQ, type of material, type of property, visual perception, complexity of task, and the like on the acquisition of conservation, and (d) criticizing the methodology of Piaget and/or offering constructive alternatives. There has, however, been little experimental examination of some of the theoretical constructs which purport to explain the ontogenesis of conservation in the minds of children. The task of examining some of these theoretical constructs in the context of experimental conditions is of fundamental importance, not only from the viewpoint of child psychology, but also from the viewpoint of the education of children. If we are to teach children to conserve properties of objects or sets of objects, we would hope to base the learning experiences we provide on theoretical constructs which have some experimental operability as well as hypothetical elegance. Accordingly, this paper (a) identifies and describes one of the basic notions thought to underlie the acquisition of conservation, and (b) reports a study which attempted to

assess the efficacy of the construct in accounting for the conservation responses of children.

The Theoretical Construct

Since Piaget is the originator of the technique of using conservation as a criterion in the assessment of concept acquisition, and since Piaget has contributed voluminous to the theory of acquisition of conservation in children, a logical source for theoretical constructs is his writing. In selecting a construct of several constructs presented by Piaget, three criteria were kept in mind. The construct had to possess properties such that: (a) it could be used to distinguish effectively between nonconservers and conservers, at least in theory, (b) it could be used to describe the developmental transition from nonconservers to conservers, and (c) it was believed by Piaget to be of fundamental importance. According to Piaget:

In physical reality there are states, and in addition to these there are transformations which lead from one state to another. . . . In children of the higher stages and in adults . . . any given state is understood to be the result of some transformation and the point of departure for another transformation. But the pre-operational child does not understand transformations. He does not have the operations necessary to understand them so he puts all the emphasis on the static quality of the states. It is because of this, for example, that in the conservation experiments he simply compares the initial state and the final state without being concerned with the transformation (Piaget, 1964, p. 20).

There are two basic constructs given in the above quote: States, and transformations. A nonconservers places all his attentions "on the static quality of the states," while a conservers takes into consideration the transformation which links one state to another. The transition from a nonconservers to a conservers is described in terms of an understanding of transformations. Thus, the notions of state and transformations satisfy the selected criteria.

States and Transformations: A Rationale

The idea of using states and transformations as a theoretical construct designed to render some structure to our thought is an idea which has been used in many different disciplines. Langer, for example, in her book on logic (1953, p. 20) uses the notions of states, and transformation from one state into another, to distinguish between our knowledge of something, and our knowledge *about* something. The former is a direct sensuous knowledge of things, the kind of knowledge a baby has of the crib, of the blankets, or of its mother's breast. However, the baby does not know *about* his mother's breast. He does not know about the relative location in space, although at times he will know of their location;

he does not know how various cultures view them, although he may know of his own reaction to them; he does not know about the effects of lactation, although he certainly may know of these effects and so on. In other words, to have knowledge *about* an object (or a state) requires more than just direct sensuous contact with it. A young school boy learning linear measurement by rote may have knowledge of measurement since he has had direct sensuous contact with rulers and other measuring apparatus, but he may have no knowledge *about* measurement, and indeed he may not even conserve length (Lovell, Healey, and Roland, 1962). To have knowledge *about* linear measurement is indeed to go far beyond the mere application of a measuring stick. Knowledge *about* linear measurement (knowing for example that length is additive, or that rotations and translations do not change length) gives meaning to knowledge of linear measurement (i.e., gives meaning to the mechanical application of a measuring stick).

In summary, the understanding of transformations which relate one state to another turns knowledge of each state into knowledge *about* the states. Operational thought could be thought of as emerging when a sufficient understanding of transformations enables the child to integrate these states into a reversible system of states, this systematization emerging out of a set of transformations linking the states, each set of transformations inducing the inference of invariants (or properties) which characterize the states being systematized, and hence give rise to conservation (and many other refinements).

For example, when a sufficient understanding of the ideas embodied in euclidean transformations enables the child to systematize space into a frame of reference or container for physical movement, then each location of a solid body becomes reversibly related with any other location assumable by that body. The set of euclidean transformations, therefore, defines certain invariants such as length, area, or shape which emerge as properties characterizing the objects and, thus, sets the basis for the conservation of these properties. Likewise, when a sufficient grasp of the basic ideas embodied in topological transformations enables the child to conceive of each state of a piece of plasticine as being merely a different form of the same basic substance, each form being transformable into any other such form, then the inducement of the inference that some basic underlying thing is remaining invariant under all such transformations sets the basis for the emergence of the conservation of substance, weight, and other properties which remain invariant.

In summary, it is the growth in knowledge *about* states, or, in other words, the growth in knowledge of the *transformations* which relate one state to another state, which transforms a nonconservator into a conservator.

As an interesting aside, it might be conjectured that it is knowledge *about* transformations which leads the child from concrete operations to formal operations.

Experimental Studies

The studies which have the most direct bearing on the problem of identifying important variables thought to underlie the acquisition of conservation are those which attempted to ascertain the efficacy of various training methods for inducing conservation in children. Of the several studies done, it is probably safe to describe only three (Beilin, 1965; Wallach and Sprott, 1964; and Sigel, Roeper and Hooper, 1966) as having experienced, to a significant degree, success in inducing conservation. However, before considering these studies, it may be pertinent to discuss a study by Greco (1959, cited in Smedslund, 1961a, and in Lauredeau and Pinard, 1962) who tested the efficacy of empirical rule learning acquired by external reinforcement, and interpreted his findings in terms of the understanding of transformations.

Greco's apparatus consisted essentially of three coloured balls attached to a rod which was pulled through a cardboard tube which could be rotated in multiples of 180 degrees. The child observed the balls as they entered the tube, and after one or two rotations, tried to predict the colour of the first ball exiting. Two groups, designated D and A, of children aged 5-10 to 7-1 were given training; group D (the empirical group) practiced the outcomes of one and two rotations *separately* to perfection, while group A learned a *mixture* of one and two rotations to perfection. A posttest given some one to three months later revealed that nearly all of group D had forgotten everything, while group A retained almost everything; furthermore, group A showed considerable but not complete generalization to "n" rotations. Greco concluded that learning occurred in two ways: by learning the outcomes of one and two rotations separately as empirical rules which have no "necessary" quality about them, and by profound internal re-structuring forced by the more complex nature of the presentation of a mixture of rotations. Greco inferred that the mixture led the subjects to search for relations between one and two rotations, and to understand that the composite transformation of two rotations is simply the product of one rotation followed by another. Thus, Greco's explanation highlights the role of the understanding of transformations (rotations in this case).

Wallach and Sprott (1964) also interpreted the success of their training method in terms of an understanding of transformations. Using a training method of inducing number conservation which (a) required the child to predict whether a one-to-one correspondence could be re-established between two sets of discrete complimentary objects (dolls

in beds) after the objects had undergone an additive-subtractive or spatial transformation or both, and (b) required the child to attempt to reconstruct the one-to-one correspondence, the investigators found that the method was successful beyond the .001 level of significance. More conclusively, a posttest given from two to three weeks later showed that the experimental group maintained their conservation. Wallach and Sprott reasoned that the crucial variable underlying conservation is "reversibility," "reversibility" being definable in terms of transformations as the following summary quotation indicates:

a property is regarded as conserved under an operation removing a defining attribute, when it is expected that the inverse operation will lead to the attribute's reappearance at the same value [p. 1069].

It should be noted that the term "operations" is used in the above context precisely the way the present writer uses the term "transformation," and parenthetically, from a mathematical point of view, any particular operation can be conceived of as being a special kind of transformation, and Piaget's theories are extensively rooted in mathematics.

However, Lovell and Ogilvie (1961), using the now classic plasticine balls experiment to test for the conservation of weight and substance, reported that 46% of the nonconservers showed evidence of being able to recall that the balls were equal before the shape transformation was applied to one of them. Defining reversibility as "the capacity to show awareness of the equality of the balls before the experiment (p. 139)," the authors concluded that reversibility is not a sufficient condition for conservation. Supposing that the discrepancies between Wallach and Sprott's conclusion and that of Lovell and Ogilvie is not accounted for by the difference between number conservation and weight conservation, then what is the explanation of these contrary findings? Wallach and Sprott suggest that mere "dispositional reversibility" (such as defined by Lovell and Ogilvie) does not require the child to actually imagine the inverse transformation which will re-establish the equality. In the terminology of the present paper, the criterion of reversibility in Lovell and Ogilvie's definition is a *state* criterion; the child has only to recall the equality state prior to the transformation. The criterion in Wallach and Sprott's definition is a *dynamic* criterion; the child must actually think of the *transformation* (or operation) which will re-instate the equality state. What Wallach and Sprott are suggesting is that a child will conserve if he realizes that every spatial transformation, indeed, any conserving transformation, necessarily implies its own inverse. They point out that conservation is not necessarily deducible from the fact that every shape transformation has an inverse, but that to *assume* that this is the case makes for cognitive economy.

Reversibility also played a part in a training study done by Sigel, Roeper, and Hooper (1966). However, the reversibility training in this

study was only one of three types of training given to all the children, the other types being "multiple classification" and "multiple relationality." It is impossible to determine which of these three aspects contributed most to the success of the training, or whether the interaction of these aspects was the crucial variable.

Another recent study in which some success was experienced in inducing number conservation in children was Beilin's (1965). In general, his procedure consisted of the familiar pretest-training-posttest design which he used on 170 children aged 5 to 11. The only training method of the four used which produced significantly greater gains than those shown by the control group was a highly verbal method Beilin called "Verbal Rule Instruction." It appears highly unlikely that "Verbal Rule Instruction" could induce any profound cognitive reorganization, and perhaps profound cognitive reorganization is not required. This, however, would contradict the findings of Piaget (1950, 1960), Wohlwill and Lowe (1962), Smedslund (1961a, b, c, d, e, f), Feigenbaum and Sulkin (1964), and Ervin (1960). Perhaps the explanation is that Beilin's posttest apparatus, materials, and procedure were almost identical to those used in the training sessions so that superficial cognitive reorganization verbally induced would be sufficient to give correct responses on the posttest. Such superficial learning would not be expected to transfer to more different contexts involving the same conservation principles, and Beilin (1965) and Beilin, Kagan, and Rabinowitz (1966) have found that indeed it does not.

In summary, it may be concluded that an understanding of transformations has played a rather important if not crucial role in the training methods that were successful in inducing conservation. It appears, therefore, that the role of transformations in the acquisition of conservation merits a closer examination. It may even be hypothesized that the most crucial variable relating to the acquisition of conservation is an understanding of transformations.

Method

In order to test the hypothesis that an understanding of transformations is an important factor in the acquisition of conservation, the investigator translated this general problem into a specific one dealing with the conservation of length. A special test of length conservation consisting of 32 items with built in euclidean transformational and state properties classified in certain basic ways was constructed on the assumption that, when such a test was factor analyzed, the hypothesis of the importance of transformations could be answered by ascertaining whether the underlying dimensions in the test could be best interpreted in terms of transformations or not.

Subjects

The sample consisted of 62 kindergarten and grade one pupils aged 5-4 to 8-0 randomly selected from a population of 196 such pupils attending schools servicing the Canadian Forces Base at Griesbach near Edmonton, Alberta. The vast majority of the fathers of the children were of the Corporal or lower rank.

Procedure

For this study, an essentially nonverbal method was used in testing for the conservation of length under the assumption that a nonverbal method would give a "purer" measure of conservation (Braine, 1959, 1964; Sawada and Nelson, 1968).

To remove the dependence on the word "length" to communicate the concept of length to the child, calipers were used to define to the child that attribute about which he was to make decisions.¹

A training session of about 12 to 15 minutes per S was given to teach the child how to use the calipers in conjunction with a three-choice response apparatus in order to facilitate S's decision-making about how a given transformation affected the length of various rods. The child was first taught to arrange the rods in a straight compact line before attempting to apply the calipers of which there were three sizes—9 cm., 9.5 cm., and 10 cm. It was arranged so that the calipers always fitted the rods in one of only three ways:

1. The calipers were .5 cm. too short for the calipers.
2. The calipers were just right for the rods.
3. The calipers were .5 cm. too long for the calipers.

Each of the three different "fits" described above was associated with one of three "doors" on the response apparatus, the association being facilitated by placing a "model fit" above each of the doors and by training S to apply a pair of calipers to a set of rods, to note how the calipers fit, to locate the equivalent model fit, and finally, to open the corresponding door which revealed a candy reward. Most S looked upon the above procedure as a game, and all 62 S could easily find the candy. The "game" was slightly modified by having S predict the fit after the calipers were removed and the rods given a transformation. The resulting procedure was as follows:

1. A rod or set of rods and a pair of calipers were placed in front of S.
2. S applied the calipers to the rods noting the fit.
3. E removed the calipers and applied a transformation to the rods.
4. S predicted how the calipers would fit if he were allowed to put them on again. "How would they fit if you tried to put them on again? Like this, or this, or this?" E asked, pointing at the model fits. "Find the candy."

¹ A fuller description of the testing technique used in the present study is given in Sawada and Nelson (1968).

Three practice items were given using steps 1 to 4 above, and all S appeared to understand what was required. The test was then given using steps 1 to 4 for each of the 32 items. A fuller description of the testing technique exists elsewhere (Sawada and Nelson, 1968).

The Test

Since the important variables under study are transformations, and states, a variety of euclidean transformations and a variety of state attributes were incorporated into the 32 items resulting in items which could be classified into several subsets according to the kind of transformation or state attribute embodied in the item. The transformations could be classified into the following kinds:

1. Translations—e.g., a wooden rod is pushed 20 cm. horizontally.
2. Rotations—e.g., a wooden rod is rotated 60° .
3. In-plane transformations—a transformation which can be executed entirely within one or two dimensions.
4. Out-plane transformations—a transformation which requires three dimensions before it can be executed: i.e., the plane must be conceived of as immersed in a three-space.

The above 4 kinds of transformations all leave length invariant. The test would be crucially incomplete without transformations which alter length. Therefore, a fifth classification was included.

5. Noninvariant transformations—a transformation which alters the length of an object.

It should be noted that for transformations 1 to 4 above, a precise mathematical rationale can be given in terms of orthogonal line reflections (see Sawada, 1966, pp. 122-25).

The state attributes were of the following kinds:

1. Multiple segmented lengths—a length exhibited by an object consisting of two or more discrete parts.
2. Single segmented length—a length exhibited by an object consisting of a continuous whole.
3. Relations—relations of less than, equal to, and greater than (F, =, J) between the rod and calipers.
4. Material—wood, plasticine, cigarette.
5. Illusions—these were variants of (a) Muller-Lyer, (b) Baldwin, (c) Opel inverted T, (d) the use of two perspective lines. These illusions were used in an attempt to make the rods appear to have increased in length.

Table 1 gives a summary of the above dimensions, and lists the test items incorporating each of the dimensions.

Hypothesis and Data Analysis

The hypothesis can be explicitly stated as follows: When a factor analysis is performed on the conservation test used in this study, the dimensions which will define the factors will be transformational dimensions rather than state dimensions.

TABLE 1
DIMENSIONS BY WHICH THE TEST ITEMS ARE CLASSIFIED

Dimensions	Items measuring the dimension
Translations	1, 2, 3, 4, 5, 6, 7, 8,
Rotations	9, 10, 11, 12, 13, 14, 15, 16,
Noninvariant	17, 18, 19, 20, 21, 22, 23, 24,
In-plane	2, 3, 5, 8, 9, 12, 14, 15,
Out-plane	1, 4, 6, 7, 10, 11, 13, 16,
Multiple Segments	3, 4, 7, 8, 11, 12, 15, 16,
Single Segments	1, 2, 5, 6, 9, 10, 13, 14,
< (less than)	5, 6, 11, 12, 13,
= (equal to)	1, 2, 3, 4, 9, 10,
> (more than)	7, 8, 14, 15, 16,
Wooden rods	17, 18,
Plasticine rods	21, 22,
Cigarettes	23, 24,
Illusions	25, 26, 27, 28, 29, 30, 31, 32,
I (=)	25, 26, 27, 28,
I (≠)	29, 30, 31, 32,

The study was designed as a factor analytic study rather than a training study because the investigator believed that it is not known just what transformational variables are the crucial ones. Instead of incorporating one or two transformational variables in a training method, the investigator included several transformational variables, systematically classified in some way, in a test of conservation, and performed a factor analysis on the test to ascertain whether understanding of transformations is indeed an important factor, and, if so, which transformations are important.

The Principal Axis method of factoring programmed for the IBM 7040 was employed using the fortran coded programs of the Division of Educational Research Services of The University of Alberta. Since the Principle Axis method extracts factors in the order of the amount of the variance accounted for, it was assumed that any dimension crucial to the performance on the test would show itself in at least one of the first few factors. Accordingly five factors were extracted and interpreted.

In order to facilitate the interpretation of the factors, the factor matrix was rotated to approximate "simple structure" using the Equimax criterion programmed for the IBM 7040. Table 2 presents the rotated matrix of factor coefficients, and Table 3 presents the transformation matrix linking the rotated matrix to the unrotated matrix.

Interpretation of the Factors

For the purposes of facilitating the identification of the factors, Table 4 was drawn up denoting the major configurations in the factor pattern. The table is based on factor coefficients, which are at least as high as .350; coefficients between .350 and .499 are designated with an "m" for "medium" factor loading, and coefficients between .500 and 1.000 are designated by an "H" for "high" factor loading.

Items 1 to 16 load on factors I and II while items 17 to 32 load on factors III, IV, and V. The column headed by "Relevant Dimension" aids in the interpretation of the factors. Factor III is essentially an I(=) factor (illusions involving equality between the rod and the calipers).

TABLE 2
ROTATED FACTOR MATRIX

Test Item	I	II	Factor III	IV	V	Communality
1	.390	.417	.191	— .124	.271	.451
2	.363	.738	— .034	— .098	.177	.719
3	.698	.468	.234	— .044	.169	.791
4	.646	.540	.218	.069	.106	.776
5	.041	.643	.297	.243	— .085	.557
6	— .083	.808	.030	.154	— .088	.691
7	.462	.398	.325	.125	— .270	.566
8	.642	.071	.061	.052	.053	.427
9	.512	.638	— .012	— .013	— .070	.675
10	.561	.461	.062	— .170	.250	.623
11	.591	.232	— .044	— .118	.317	.519
12	.804	.075	— .047	.164	— .107	.692
13	.595	— .053	.452	— .035	.120	.572
14	.139	.714	.100	— .125	— .031	.556
15	.735	.400	.118	.064	— .101	.728
16	.671	.194	— .036	— .008	— .083	.496
17	— .296	.104	.212	.225	.483	.427
18	— .087	.005	.024	.339	.197	.162
19	— .017	.130	.140	.611	.071	.415
20	.207	.182	— .186	.307	.414	.377
21	— .059	— .030	— .125	.190	.759	.632
22	— .100	— .130	— .331	.388	— .130	.304
23	.027	— .221	.164	.309	.352	.296
24	.100	— .082	.294	— .072	.675	.564
25	— .054	.073	.705	.147	— .003	.527
26	.233	— .138	.551	— .124	.351	.515
27	— .035	.114	.752	— .001	.186	.614
28	.142	.180	.730	— .034	— .088	.594
29	— .006	— .065	.262	.494	— .015	.318
30	.212	.222	— .142	.402	.111	.289
31	.078	— .042	— .065	.755	.068	.587
32	— .059	— .011	— .131	.659	.111	.467
Variance accounted for	5.005	4.095	2.923	2.596	2.259	16.927
% of Common Variance	29.86	24.19	17.27	15.34	13.34	100.00

TABLE 3
TRANSFORMATION MATRIX

-.735	-.613	-.268	-.032	-.108
-.182	-.136	.503	.593	.587
-.129	-.122	.706	-.686	.019
-.550	.624	.236	.222	-.452
-.329	.449	-.349	-.359	.663

Note: If T is the transformation matrix and if U is the unrotated matrix and R the rotated matrix, then $UT = R$.

TABLE 4
'SCHEMATIC' FACTOR MATRIX

Test item	Relevant Dimensions			Factor					Test item
				I	II	III	IV	V	
1	Single,	=	m*	m					1
2	Single,	=	m	H					2
3	Multi,	=	H	m					3
4	Multi,	=	H	H					4
5	Single,	<		H					5
6	Single,	<		H					6
7	Multi,	>	m	m					7
8	Multi,	>	H						8
9	Single,	=	H	H					9
10	Single,	=	H	m					10
11	Multi,	<	H						11
12	Multi,	<	H						12
13	Single,	<	H			m			13
14	Single,	>		H					14
15	Multi,	>	H	m					15
16	Multi,	>	H						16
17	Non-inv							m	17
18	Non-inv								18
19	Non-inv						H		19
20	Non-inv							m	20
21	Non-inv							H	21
22	Non-inv						m		22
23	Non-inv							m	23
24	Non-inv							H	24
25	I(=)					H			25
26	I(=)					H			26
27	I(=)					H			27
28	I(=)					H			28
29	I(≠)						m		29
30	I(≠)						m		30
31	I(≠)						H		31
32	I(≠)						H		32

*"m" designates a "medium" factor loading (.350 - .499)
"H" designates a "high" factor loading (.500 - 1.000)

Factor V is essentially a noninvariant transformation factor. However, the identity of factors I and II is by no means obvious. By using the dimensions single and multiple segmented lengths, and the relations $<$, $=$, $>$, some identity can be given to factors I and II. The primary dimensions seem to be single and multiple segmented lengths, with factor I being a multiple segment factor, and factor II being a single segment factor. However, this dichotomy is modified by the relations $<$, $=$, $>$. When the relations is $<$, the items tend to load higher on factor I. When the relations is $>$, the items tend to load higher on factor II. When the relations is $=$, the items tend to equalize on both factors. These remarks on the interpretation of factors I and II can be verified in general by examining Table 4.

For the purposes of summary, it should be noted that all items loading on factors I and II are items which leave length invariant. Thus, if factors I and II are grouped together, they can be thought of as embodying the notion of "invariant transformations" and together can be contrasted with factor IV which may be called "noninvariant transformations." Hence, the major transformational variable in the test can be represented by the classification of transformations into those which leave length invariant, and those which do not. A finer classification of transformation does not seem justified on the basis of the above analysis. The state variables $I(=)$ and $I(\neq)$, defining factors III and IV respectively, emerge as the major state variables; single segments, multiple segments and the $<$, $=$, $>$ relations emerge as secondary state variables, defining as they do, the distinction between factors I and II.

Discussion

The results indicate that understanding transformations is of no more importance than the understanding of states as concerns the acquisition of the conservation of length. The fact that state dimensions contributed as much to the identification of the factors as did the transformational dimensions suggests that training studies designed to teach children to acquire conservation should incorporate in their training sessions mental operations which explicitly incorporate both state and transformational dimensions. Moreover, transformations which change the magnitude of the attribute must be included as well as transformations which leave the magnitude of the attribute invariant.

The Sigel, Roeper, Hooper study (1966), which was successful in training children in acquiring conservation, used a training method which was designed to increase the child's awareness and facility with the mental operations associated with: multiple classification, multiple relationality, and reversibility. Multiple classification and multiple relationality (as described by Sigel, Roeper, and Hooper) are dependent essentially

on state attributes whereas reversibility is dependent essentially on understanding physical transformations. It appears then that Sigel and his colleagues included both aspects in their training. Had they left out reversibility, their training method may not have been nearly as successful as it was.

This study concludes with the inference that the understanding of state attributes and of transformations from one magnitude of the attribute to another are equally important to the acquisition of conservation.

Moreover, this inference appears likely from a theoretical point of view. Just as transformations give meaning to the states transformed, knowledge of state attributes gives meaning to the transformations applied. Without direct knowledge of states, the effect of the transformation on the states becomes indeterminant. With meaningless states, there can only be meaningless transformations. With meaningless transformations there can only be rote knowledge of states. As Piaget might say, first the child must differentiate reality into attributes of states, and transformations from one magnitude of the attribute into another; then he must co-ordinate these in an operational system of states and transformations.

References

- Beilin, H. Learning and operational convergence in logical thought development. *Journal of Experimental Psychology*, 1965, Vol. 2, 317-339.
- Beilin, H., Kagan, J., and Rabinowitz, R. Effects of verbal and perceptual training of water level representation. *Child Development*, 1966, Vol. 37, 317-329.
- Braine, M. D. S. The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychological Monographs*, 1959, Vol. 73, (Whole No. 475).
- Braine, M. D. S. Development of a grasp of transitivity of length: A reply to Smedslund. *Child Development*, 1964, Vol. 35, 799-810.
- Ervin, S. M. Training and logical operation by children. *Child Development*, 1960, Vol. 31, 555-564.
- Feigenbaum, K. E., and Sulkin, H. Piaget's problem of conservation of discontinuous quantities: A teaching experience. *Journal of Genetic Psychology*, 1964, Vol. 105, 91-97.
- Langer, S. M. *An Introduction to symbolic logic*. New York: Dover, 1962.
- Laurendeau, M., and Pinard, A. *Casual thinking in the child*. New York: International Universities Press, 1962.
- Lovell, K., Healey, D., and Roland, A. D. The growth of some geometrical concepts. *Child Development*, 1962, Vol. 30, 751-767.
- Lovell, K., and Ogilvie, E. A study of the conservation of substance in the junior school child. *British Journal of Educational Psychology*, 1960, Vol. 30, 109-118.
- Piaget, J. *The psychology of intelligence*. London: Routledge & Kegan, Paul, 1950.
- Piaget, J. *The child's conception of geometry*. New York: Harper & Row, 1960.
- Piaget, J. Development and learning. In R. Ripple and V. Rockcastle (Eds.), *Piaget rediscovered: A report of the conference on cognitive studies and curriculum development*. Cornell University, 1964. Pp. 7-20 (mimeo).

- Sawada, D. Transformations and concept attainment: A study of length conservation in children. Unpublished Masters Thesis, The University of Alberta, 1966.
- Sawada, D., and Nelson, L. D. Conservation of length: Methodological considerations. *Alberta Journal of Educational Research*, 1968, Vol. XIV, No. 1.
- Sigel, I. E., Roeper, A., and Hooper, F. H. A training procedure for acquisition of Piaget's conservation of quantity: A pilot study and its replication. *British Journal of Educational Psychology*, 1966, Vol. 36, 301-311.
- Smedslund, J. The acquisition of conservation of substance and weight in children: I. Introduction. *Scandinavian Journal of Psychology*, 1961, Vol. 2, 11-20. (a)
- Smedslund, J. The acquisition of conservation of substance and weight in children: II. External reinforcement of conservation of weight and of the operations of addition and subtraction. *Scandinavian Journal of Psychology*, 1961, Vol. 2, 71-84. (b)
- Smedslund, J. The acquisition of conservation of substance and weight in children: III. Extinction of conservation of weight acquired "normally" and by means of empirical controls on a balance scale. *Scandinavian Journal of Psychology*, 1961, Vol. 2, 85-87. (c)
- Smedslund, J. The acquisition of conservation of substance and weight in children: IV. An attempt at extinction of the visual components of the weight concept. *Scandinavian Journal of Psychology*, 1961, Vol. 1, 153-155. (d)
- Smedslund, J. The acquisition of conservation of substance and weight in children: V. Practice in conflict situations without external reinforcement. *Scandinavian Journal of Psychology*, 1961, Vol. 2, 156-160. (e)
- Smedslund, J. The acquisition of conservation of substance and weight in children: VI. Practice on continuous versus discontinuous material in conflict-situations without external reinforcement. *Scandinavian Journal of Psychology*, 1961, Vol. 2, 203-210. (f)
- Wallach, and Sprott, R. Inducing number conservation in children. *Child Development*, 1964, Vol. 34, 1057-1071.
- Wohlwill, J. F., and Lowe, R. C. Experimental analysis of the development of conservation of number. *Child Development*, 1962, Vol. 33, 153-167.

BOOK REVIEW

THE SPECIFICATION AND MEASUREMENT OF LEARNING OUTCOMES

By: David A. Payne (Waltham, Massachusetts: Blaisdell, 1968).

The author states that the *Specification and Measurement of Learning Outcomes* is intended for:

1. Classroom teachers,
2. Classes in test construction at both undergraduate and graduate levels,
3. Education and psychology classes requiring an overview of measurement.

The ten chapters of the book may be divided into four broad sections. The first section, chapters one to five, after a brief orientation, deals with test planning and construction. The interrelationship of educational objectives, teaching, and measurement-evaluation activities is brought out. Next, procedures for developing a test are presented, beginning with the table of specifications and then leading into types of instruments, item writing, and the use of essay tests.

The second section, chapters six to eight, deals with deriving and applying test data. The section begins with statistical procedures, then proceeds to test validity and reliability, and finally, to item analysis by hand methods.

The third and fourth sections consisting of one chapter each deal respectively with the use of standardized achievement tests and assigning grades. The approach used by the author is one of dealing with the problems commonly found in the two areas. Such problems as choosing an achievement test, sources of pertinent information, misconceptions regarding achievement tests, are discussed in chapter nine. In the final chapter, the more important problems discussed are: the relative or absolute standard, achievement and effort as bases for marks, combining scores, and adjusting grades.

The book suffers from over-summarizing of complex material. This is to be expected in view of the fact that the author tried to cover in about 200 pages most of what is found in much longer measurement texts plus other sources such as Bloom's and Krathwohl's "Taxonomies." Tabular summaries of such things as "principles of oral examinations" and "critical questions useful in reviewing educational achievement test items" do not seem to promote learning. Also, the chapter on statistical

procedures seems to expect too much of the reader if he is, indeed, part of the audience for whom the book is intended. A student or teacher with no knowledge of statistics, instead of acquiring some knowledge, will likely fall victim to "symbol shock."

The chapter dealing with item analysis was disappointing to this reviewer. The "show of hands" technique is recommended for small groups of pupils and a brief description is given. However, there is no suggestion that machine-based methods are becoming increasingly available to teachers. This seems to be an oversight, in that teachers more commonly deal with several sections of a course and the groups are therefore too large to be involved in the simple hand methods. It must be conceded that this viewpoint represents a personal bias.

It is possible that the book would be of value to classroom teachers seeking to gain a quick review of the field of educational measurement, especially if it were used in a workshop situation. Students in courses which have a unit on measurement might also find the book of value; however, it is likely that excerpts from well known texts such as Ebel's, Thorndike and Hagen's, or Downie's would be more valuable. The book would not be of much value as a text or primary reference for undergraduate or graduate measurement courses.

It is obvious that the author delved deeply into the writings on educational measurement. The suggested readings found at the end of each chapter, and the list of references at the end of the book, provide a valuable summary of the literature in the field.

V. R. NYBERG

The University of Alberta

NOTES ON CONTRIBUTORS

DALE ANDERSEN is an Assistant Professor of Education at Washington State University, Pullman, Washington.

RAM K. GUPTA is an Assistant Professor in the Department of Educational Psychology, The University of Alberta. He holds B.Sc., L.T., and M.Ed. degrees from India; and M.A. from London, and a Ph.D. from Minnesota. He has been employed on a number of research projects in Minnesota and has published in a number of journals. His special interests are in measurement theory, research design, and statistics.

H. D. HEMPHILL, B.A., M.Ed., is a graduate of the Universities of British Columbia and Alberta. He is currently completing doctoral requirements at The University of Alberta, Department of Educational Administration. He is the author of articles on the teaching of modern mathematics, educational finance and the educational use of automatic data processing.

R. S. MACARTHUR, whose doctoral study was carried out at the University of London, has for some years been a Professor of Educational Psychology at The University of Alberta. His current research interests are on the theme of cognitive development in cross-cultural context. His numerous publications include reports of work with native peoples of the Canadian North and of Central Africa.

L. D. NELSON obtained his Ph.D. from the University of Minnesota after graduating from The University of Alberta. He is co-editor of MATHEX published by Encyclopedia Britannica of Canada. He is a member of the Publications Committee and a member of the Editorial Board for a yearbook on the history of mathematics, National Council Teachers of Mathematics.

VERNER R. NYBERG is an Associate Professor at The University of Alberta. He holds B.Sc., B.Ed., and M.Ed. degrees from The University of Alberta and an Ed.D. from U.C.L.A. His special interests are psychological measurement and construction of classroom tests.

AUBREY H. RODEN studied at The University of California, Berkeley and at The University of Alberta, Canada. He has published articles in journals such as the *British Journal of Psychology*.

DAIYO SAWADA holds a joint appointment as an Assistant Professor in Elementary Education and in the Division of Educational Research Services at The University of Alberta. He has had articles published in the *Mathematics Teacher* and the *Alberta Journal of Educational Research*.

MALCOLM J. SLAKTER studied at Syracuse University and at The University of California, Berkeley. He has had articles published in journals such as *Biometrika* and the *American Educational Research Journal*.

HERBERT E. STILES received his Master of Education degree from The University of Montana. He is now the principal of Didsbury High School, Didsbury, Alberta.



**"Edmonton's
Most Complete
Book Store"**

**101 Street
Just Off Jasper Ave.**

**Phone
422-0507 or 422-0556
EMPIRE BUILDING —
10117 - 101 Street**

THE BOOKMARK LTD.

Through an arrangement with New York University Press, publications of The National Foundation for Educational Research in England and Wales will be available in Canada from Copp Clark.

INCLUDED ARE:

THE i.t.a. SYMPOSIUM

Research Report on the British Experiment

by **JOHN DOWNING**

This is the first definitive report on the famous initial teaching alphabet (i.t.a.). Based on a study of 4,000 children, half using i.t.a. and half using traditional orthography, it offers careful comparisons of methods, results, and conclusions. Included are a number of short papers by prominent reading specialists.

\$4.50

STANDARDS AND PROGRESS IN READING

by **JOYCE M. MORRIS**

This longitudinal report examines the first comprehensive British research into factors influencing children's achievements in reading. These factors include the child himself, the home, and the school, and assesses the attitudes and methods of teachers. Case studies are included.

\$7.50

CONCEPT GROWTH AND THE EDUCATION OF THE CHILD

by **J. G. WALLACE**

A comprehensive survey on the present state of our knowledge in the study of the growth and development of concepts. The first section deals with research in the traditional behaviour theory and the second is mainly devoted to the work of Piaget and his associates.

\$6.00

THE COPP CLARK PUBLISHING COMPANY

**517 Wellington Street West
TORONTO 2B, ONTARIO**